Εθνικο Μετσοβιο Πολυτεχνειο
Σχολη Ηλεκτρολογων Μηχανικων και Μηχανικων Υπολογιστων
Τομεας Τεχνολογιας Πληροφορικης και Υπολογιστων

Ανάλυση και Στατιστικά Μοντέλα για το Δίκτυο Συναλλαγών του Ευρωπαϊκού Μηχανισμού Εμπορίας Δικαιωμάτων Εκπομπής Αερίων Ρύπων (EU ETS)

Διπλωματικη Εργασια

του

ΓΕΩΡΓΙΟΥ ΡΟΥΣΣΑΚΗ

Επιβλέπων: Δημήτριος Φωτάκης
Αναπληρωτής Καθηγητής Ε.Μ.Π.

Εργαστηριο Λογικης και Επιστημης Υπολογιστων
Αθήνα, Οκτώβριος 2021

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών
Τομέας Τεχνολογίας Πληροφορικής και Υπολογιστών

Ανάλυση και Στατιστικά Μοντέλα για το Δίκτυο Συναλλαγών του Ευρωπαϊκού Μηχανισμού Εμπορίας Δικαιωμάτων Εκπομπής Αερίων Ρύπων (EU ETS)

Διπλωματική Εργασία

του

ΓΕΩΡΓΙΟΥ ΡΟΥΣΣΑΚΗ

Επιβλέπων: Δημήτριος Φωτάκης
Αναπληρωτής Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 11/11/2021.

(Υπογραφή)	*(Υπογραφή)*	*(Υπογραφή)*
....	....	....
Δημήτριος Φωτάκης	Αριστείδης Παγουρτζής	Ευάγγελος Μαρκάκης
Αν. Καθηγητής Ε.Μ.Π.	Καθηγητής Ε.Μ.Π.	Αν. Καθηγητής Ο.Π.Α.

Εργαστηριο Λογικης και Επιστημης Υπολογιστων
Αθήνα, Νοέμβριος 2021

(Υπογραφή)

....

Γιωργος Ρουσσακης
Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών
Τομέας Τεχνολογίας Πληροφορικής και Υπολογιστών

Ευχαριστίες

Θα ήθελα να ευχαριστήσω ιδιαίτερα τον επιβλέποντα καθηγητή μου κ. Φωτάκη αρχικά για την ευκαιρία για συνεργασία που μου παρείχε και την καθοδήγηση του σε όλη τη διάρκεια της διπλωματικής, καθώς και για τη διδασκαλία του στα προπτυχιακά μου χρόνια.

Σημαντική ήταν η συνεισφορά του Σωτήρη Δήμου, του οποίου η υποστήριξη και συνεργασία ήταν καταλυτικές στην διεκπεραίωση αυτής της διπλωματικής.

Τέλος, θέλω να ευχαριστήσω την οικογένεια μου και τους φίλους μου που με στήριξαν σε όλο αυτό το διάστημα και έκαναν αυτή την πορεία εφικτή.

Περίληψη

Στην αναζήτησή λύσεων στο πρόβλημα της κλιματικής αλλαγής, πολλές κυβερνήσεις και οργανισμοί έχουν εφαρμόσει συστήματα εμπορίας ρύπων ώστε να περιορίσουν τις εκπομπές ρύπων από τον βιομηχανικό τομέα. Ένα από αυτά τα συστήματα (EU ETS) που εφαρμόζεται στα μέλη της Ευρωπαϊκής Ένωσης αποτελεί το αντικείμενο αυτής της διπλωματικής.

Παρουσιάζουμε το EU ETS(European Emission Transaction System) ως ένα δίκτυο συναλλαγών και εξετάζουμε τα χαρακτηριστικά που αποτελούν την κύρια δομή του. Δείχνουμε ότι η τοποθεσία επηρεάζει σημαντικά την επιλογή των συμμετεχόντων αναφορικά με τις συναλλαγές τους, καθώς η δομή του δικτύου υποδεικνύει ότι οι κόμβοι μεγάλων βαθμών συναλλάσσονται κυρίως με κόμβους χαμηλότερων βαθμών, γεγονός που επιβεβαιώνουν και οι εικόνες του δικτύου. Παρά τις αλλαγές μεταξύ φάσεων, η βασική δομή και τα χαρακτηριστικά του γράφου παραμένουν σταθερά επί το πλείστον και δείχνουν να έχουν ετήσια περιοδικότητα. Στην αναζήτηση του πυρήνα του δικτύου, συναντάμε τους ίδιους τύπους εταιριών να έχουν την υψηλότερη κεντρικότητα και όπως αποδεικνύεται, εκείνους που έχουν ανταλλάξει συνολικά το μεγαλύτερο όγκο αδειών. Κατά τη δοκιμή υποθέσεων για το δίκτυο, αποδεικνύουμε ότι παρουσιάζει πύκνωση κατά την πάροδο του χρόνου και μείωση της (ουσιαστικής) διαμέτρου του. Το δίκτυο επίσης παρουσιάζει συμπεριφορά στην οποία ¨οι πλούσιοι γίνονται πλουσιότεροι' καθώς αποδεικνύεται ότι ακολουθεί scale-free κατανομή βαθμών.

Έπειτα, χρησιμοποιούμε γνωστές μετρικές για συγκρίσεις δικτύων και παρέχουμε μια ακριβή απεικόνιση ομοιότητας του κάθε μοντέλου με το πραγματικό δίκτυο αναφορικά με τη δομή. Στη συνέχεια, περιγράφουμε κάποια μοντέλα που πιστεύουμε ότι μπορούν να αναπαράγουν τη δομή του δικτύου και δείχνουμε πόσο κοντά τα τεχνητά δίκτυα πλησιάζουν στο πραγματικό. Για την εκδοχή χωρίς κατευθύνσεις ακμών, τροποποιούμε το μοντέλο Barabasi Albert ώστε να δέχεται μη ακέραιες τιμές το οποίο αποδεικνύεται η πιο κοντινή αναπαραγωγή του πραγματικού δικτύου στις πιο σταθερές φάσεις. Στην εκδοχή με κατευθύνσεις ακμών, το Community Guided Agreement δείχνει να μπορεί να αναπαράγει ικανοποιητικά σημαντικά γνωρίσματα του δικτύου.

Λέξεις Κλειδιά

Ανταλλαγή Ρύπων, Συστήματα Ανταλλαγής και Εμπορίας Ποσοστώσεων, EU ETS, Ανάλυση Δικτύων, Σύγκριση Δικτύων, Στοχαστικά Μοντέλα Δικτύων

Abstract

Seeking solutions to climate change, many governments and organizations have implemented emission trading systems to limit emissions from the industrial sector. One of these systems(EU ETS enforced on and by the members of the EU is the subject of this thesis.

We present the EU ETS(European Emission Transaction System) as a transaction network and examine the structural characteristics that comprise the main structure of the network. We show that location affects greatly the choice of actors to transact while the structure of the network indicates that the larger degree nodes transact mostly with lower degree nodes, something which network images corroborate. Despite the changes between phases, the basic structure and characteristics of the graph remain largely constant and appear to have an annual periodicity. In the search for the core of network, we find the same types with highest centrality and as we prove, the ones with the largest volume of traded allowances. When testing network hypotheses, we prove that the network exhibits densification over time and shrinking of the (effective) diameter. The network is found to exhibit rich-get-richer behavior as it is proven it follows scale-free degree distribution.

Afterwards, we make use of known measures for network comparison, in order to provide an accurate depiction of the similarity each model exhibits to the real network in terms of structure. Following that, we describe some models that we believe can recreate the structure of the network and we show how closely the artificial networks resemble the real one. For the undirected version, we modify the Barabasi-Albert to include non-integer values which proves to be the closest recreation of real network during the more stable phases. In the directed case, the Community Guided Agreement appears to reproduce the important characteristics of the network more accurately.

Keywords

Emissions Trading, Cap-and-Trade Systems, EU ETS, Network Analysis, Network Comparison, Stochastic Network Models

Περιεχόμενα

Κατάλογος Σχημάτων

Κατάλογος Πινάκων

Κεφάλαιο 1

Εκτεταμένη Περίληψη

1.1 Σύστημα Εμπορίας Ρύπων

Τον τελευταίο μισό αιώνα, η κλιματική αλλαγή αποτελεί ένα σημαντικό πρόβλημα που μαστίζει την ανθρωπότητα. Η απάντηση της παγκόσμιας κοινότητας ήταν η δημιουργία του Intergovernmental Panelon Climate Change(IPCC) το 1988. Η εξέλιξη του ήταν το United NationsFramework Convention on Climate Change που καθιερώθηκε το 1992 και στο οποίο συμμετέχουν 154 έθνη. Η δέσμευση των υπογεγραφόντων εθνών είναι η μείωση της ατμοσφαιρικής συγκέντρωσης των αεριών του θερμοκηπίου με σκοπό τη αποτροπή της επικίνδυνης ανθρωπογενούς επιρροής στο κλιματικό σύστημα της Γης. Η Ευρωπαϊκή Ένωση, για την ενίσχυση των στόχων των σχετικών συνόδων, δημιούργησε το σύστημα εμπορίας δικαιωμάτων εκπομπής (Emissions Trading System) , ένα cap and trade system που βάζει άνω όριο στο σύνολο των εκπομπών από τους συμμετέχοντες στο σύστημα. Στην αρχή κάθε περιόδου κατανέμενονται άδειες στις εταιρίες αναλόγως τις εκτιμώμενες εκπομπές αερίων τους. Στο τέλος κάθε περιόδου, κάθε εταιρία είναι υποχρεωμένη να πληρώσει άδειες ανάλογες με τους ρύπους που παρήγαγε. Αν παράξει παραπάνω ρύπους από τις αντίστοιχες άδειες, αναγκάζεται να αναζητήσει παραπάνω άδειες στην αγορά, διαφορετικά θα υποβληθεί σε υψηλά πρόστιμα από την Ευρωπαϊκή Ένωση. Στην περίπτωση που καταφέρει να μειώσει τους ρύπους της επαρκώς και παράξει λιγότερους ρύπους από όσες άδειες της έχουν κατανεμηθεί, μπορεί είτε να τις αποθηκεύσει για χρήση σε μετέπειτα χρονική στιγμή(με κάποιους περιορισμούς) είτε να της πουλήσει στην αγορά αδειών.

Στο σύστημα, αναγράφονται περίπου 41 τύποι συναλλαγών, με τους τύπους αυτούς να διαφέρουν ως προς τοποθεσία, λειτουργία, ομάδα, κλπ. Κάθε τύπος συναλλαγών αποτελείται από δύο αριθμούς, με τον πρώτο να υποδεικνύει τον κύριο τύπο και τον δεύτερο τον δευτερεύον. Από αυτούς τους τύπους, στο μεγαλύτερο κομμάτι της διπλωματικής θα εστιάσουμε στους 10-0, 3-0, 3-21 που αφορούν κυρίως εμπορικές συναλλαγές. Η επιλογή αυτή έγινε καθώς οι κυβερνητικές συναλλαγές όπως διανομή και συλλογή αδειών δεν αποτελούν επιλογή των οντοτήτων του δικτύου και δεν παρέχουν σημαντική πληροφορία για τη συμπεριφορά τους.

Στο σύστημα, παρατηρούμε τρεις κατηγορίες οντοτήτων, εκείνες που έχουν τη νομική υποχρέωση να παραδώσουν άδειες με βάση τις εκπομπές ρύπων τους, ονομάζονται regulated , εκείνες που είναι υπεύθυνες για το διαμοιρασμό και συλλογή αδειών, λεγόμενες governmental και οι υπόλοιπες που δεν έχουν υποχρεώσεις στο σύστημα και κυρίως επιθυμούν να βγάλουν κέρδους ως χρηματοοικονομικές οντότητες στο σύστημα, τις οποίες θα αποκαλούμε financial.

1.2 Σκοπός της Διπλωματικής Εργασίας

Το EU ETS έχει αναλυθεί από πολλές διαφορετικές οπτικές, όμως με την συνθήκη της μη δημοσίευσης συναλλαγών μέχρι 3 χρόνια αργότερα, υπάρχει σοβαρή έλλειψη δικτυακής ανάλυσης του συστήματος. Επίσης ενώ υπάρχει εκτενής βιβλιογραφία στην τιμή των αδειών, την επίδραση της οικονομικής ύφεσης στις εκπομπές του Ευρωπαϊκού τομέα ισχύος [16] , τις κινητήριες δυνάμεις της τιμής [14] και πολλά ακόμα papers σχετικά με την τιμή των αδειών, υπάρχουν λίγα που παρέχουν μοντέλα για τη δομή του δικτύου. Η πιο ολοκληρωμένη μελέτη έγινε από Karpf et al Karpf [27] το 2018. Σε αυτή τη διπλωματική παρέχουμε μια πλήρη εξερεύνηση των χαρακτηριστικών από την πλευρά του δικτύου του συστήματος και σε όση βιβλιογραφία κοιτάξαμε δεν έχουμε συναντήσει άλλη τέτοια μελέτη να γίνεται στο EU ETS και λίγες μελέτες έχουν καλύψει τόσες περιοχές. Η ανάλυση μας παρέχει μια μεθοδολογία για την κατανόηση των δομικών χαρακτηριστικών μεγάλων δικτύων και σχέσεων μεταξύ διαφορετικών τύπων κόμβων.

Στη μελέτη μας, φτάνουμε σε μια ξεκάθαρη εικόνα των προτιμήσεων για κάθε κατηγορία των κόμβων του δικτύου και αποκτάμε γνώση που δεν έχουμε δει σε λεπτομέρεια. Ένα στοιχειώδες κομμάτι της δομής είναι η συσχέτιση μεταξύ βαθμού κάθε κόμβου και την τάση να συνδεθεί με άλλους κόμβους με βάση το βαθμό τους. Στη βιβλιογραφία των κοινωνικών δικτύων, συνηθίζεται οι κόμβοι μεγάλων βαθμών να συνδέονται με αντίστοιχα μεγάλους κόμβους. Φτάνουμε στο απροσδόκητο αποτέλεσμα, ότι οι μεγάλοι κόμβοι σχηματίζουν κοινότητες με κόμβους μικρότερων βαθμών.

Αποδεικνύουμε την επαναληψιμότητα των patterns των μηνών με στατιστικές μετρήσεις που συνδράμουν με τη a priori γνώση που είχαμε για το σύστημα.

Μια συχνή περιοχή δικτυακής ανάλυσης αποτελεί η κατανομή βαθμών, στην οποία το EU ETS αποδεικνύεται ότι ακολουθεί power-law κατανομή. Παρατηρώντας τα ευρύματα των Leskovec et al [?] οι οποίοι επιχείρησαν να μοντελοποιήσουν δίκτυα χρησιμοποιώντας μη συμβατικές μεθόδους για το φαινόμενο ¨πλούσιοι γίνονται πλουσιότεροι¨, αποδεικνύουμε ότι η ουσιαστική διάμετρος του δικτύου μειώνεται με το χρόνο και υπάρχει παράμετρος α τέτοια ώστε ο αριθμός των ακμών να ακολουθεί power-law κατανομή με τον αριθμό των κόμβων με το α ως εκθετικό. Δεν έχουμε παρατηρήσει παρόμοια διερεύνηση τέτοιων υποθέσεων για το EU ETS .

Ενώ πολλές μελέτες σε κάποιο σημείο καταλήγουν στη μοντελοποίηση δικτύου, έχουμε παρατηρήσει λίγες περιστάσεις όπου το δίκτυο που δημιουργείται χρησιμοποιώντας κάποια στοχαστική διαδικασία, συγκρίνεται κάποιο δικτυακό μέτρο σύγκρισης. Σε αυτή τη διπλωματική, χρησιμοποιώντας την Απόκλιση Πορτραίτων που αποτελεί χρήσιμο εργαλείο για σύγκριση δομικών στοιχείων μεταξύ δικτύων, δείχνουμε συγκριτικά αποτελέσματα για μοντέλα, που αποτελούν λογικό επακόλουθο της ανάλυσης που έχει προηγηθεί. Τα αποτελέσματα δείχνουν ότι στην περίπτωση χωρίς κατεύθυνση ακμών, το δίκτυο μπορεί να προσεγγιστεί καλύτερα, στις πιο σταθερές φάσεις, από το Barabasi Albert μοντέλο [5] όταν το τροποποιήσουμε για να συμπεριλάβει μη ακέραιες τιμές στην παράμετρο m, ενώ στην κατευθυνόμενη περίπτωση, η κατευθυνόμενη εκδοχή του Community Guided Agreement παράγει τα καλύτερα αποτελέσματα Απόκλισης Πορτραίτων σε σύγκριση με το]Forest Fire[37] και το μοντέλο του Price [47].

1.3 Ανάλυση Δικτύου

1.3.1 Ομοφιλία Βαθμών

Ο όρος assortativity αναφέρεται σε ομοιότητα, αναφορικά με κάποιο attribute , των κόμβων. Εδώ θα ασχοληθούμε με το assortativity mixing στους βαθμούς, δηλαδή πώς οι κόμβοι όμοιου-βαθμού παρουσιάζουν την τάση να συνδεθούν μεταξύ τους. Στη βιβλιογραφία υπάρχουν διαφορετικά μοντέλα για το assortativity και εμείς θα ασχοληθούμε με το συντελεστή συσχέτισης βαθμών του Pearson μεταξύ ζευγών συνδεδεμένων κόμβων, όπως περιγράφει ο Newman[42] .

Ορίζουμε τον πίνακα e όπου η ποσότητα e_{ij} είναι το ποσοστό των ακμών του δικτύου που συνδέουν ένα κόμβο τύπου i με ένα τύπου j. Πρέπει να ικανοποιεί τους παρακάτω κανόνες αθροισμάτων.

$$\sum_{ij} e_{ij} = 1 \; \sum_j e_{ij} = a_i \; \sum_i e_{ij} = b_i$$

$$r = \frac{\sum_i e_{ii} - \sum_i a_i b_i}{1 - \sum_i a_i b_i} = \frac{Tr\boldsymbol{e} - \parallel \boldsymbol{e}^2 \parallel}{1 - \parallel \boldsymbol{e}^2 \parallel}$$

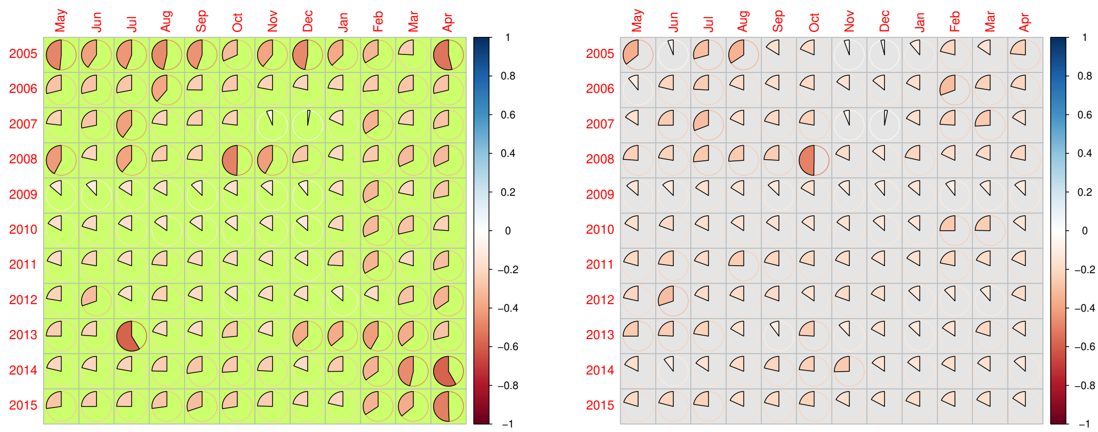
όπου e είναι ο πίνακας του οποίου στοιχεία είναι τα e_{ij} [42]

Ο παραπάνω πίνακας δ λαμβάνει υπόψη τον αριθμό των μελών σε κάθε κατηγορία. Για να αντιμετωπίσουμε το πρόβλημα παρέχουμε τον πίνακα assorativity :

$$AM = e - S$$

οπου S είναι ένας τετραγωνικός πίνακας αποτελούμενος από στοιχεία $S_{ij} = \frac{\#type_i * \#type_j}{\#edges^2}$ Ο πίνακας S αποτελεί τις τυχαίες συνδέσεις που θα συνέβαιναν με τον αριθμό των κόμβων σε κάθε κατηγορία. Έτσι ο assorativity πίνακας αντιπροσωπεύει την προτίμηση που κάθε κατηγορία δείχνει και τα στοιχεία παίρνουν τιμές στο $(-1, 1)$.

Εξετάζοντας το δίκτυο ως προς το assorativity για όλες τις συναλλαγές και κρατώντας τους τύπους συναλλαγών 10-0,3-0,3-21 παίρνουμε τα ακόλουθα αποτελέσματα



(α′) Συνολο Συναλλαγών

(β′) Συναλλαγές 10-0,3-0,3-21

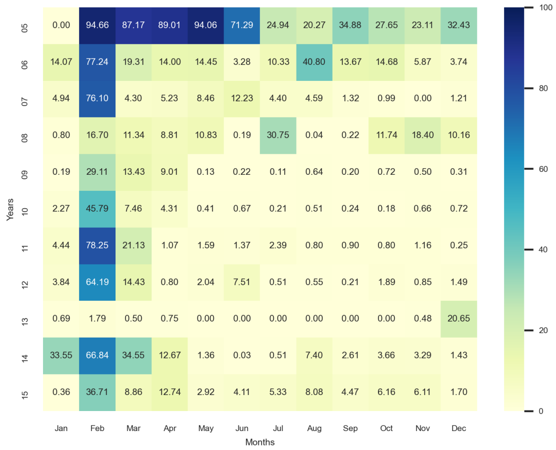
Σχήμα 1.1: Degree assortativity

1.3.2 Περιοδικότητα

Το κύριο μέρος του μηχανισμού EU ETS έγκειται στην κατανομή και συλλογή αδειών για κάθε οντότητα του συστήματος με την ποσότητα να εξαρτάται από τις εκπομπές που παράγουν μέχρι το τέλος του Απριλίου κάθε χρόνο. Η έκδοση αδειών σε regulated οντότητες γίνεται στις 28 Φεβρουαρίου σύμφωνα με το ETS handbook . Η σταθερότητα αυτών των φαινομένων υποδεικνύει επαναληψιμότητα και στους υπόλοιπους μήνες με δυνατότητα μοντελοποίησης σε ετήσια βάση.

Στους πίνακες που ακολουθούν αναγράφεται, οι συναλλαγές που έχουν τύπο 10-53 και 10-36 αποτελούν την πλειοψηφία των συναλλαγών κατανομής σε regulated οντότητες με βάση τις GhG εκπομπές, ενώ εκείνες με τύπο 10-2 αφορούν την παράδοση αδειών στο τέλος του Απρίλη.

Όπως φαίνεται, με εξαίρεση την τυχαιότητα του πρώτου έτους λειτουργίας του ETS και το 2008 που αποτελεί ανωμαλία σε όποιο χαρακτηριστικό του δικτύου μελετήσαμε, επιβεβαιώνεται η δήλωση στο handbook αν και παρατηρούμε ελαστικότητα στην έκδοση με περιορισμένο αριθμό και στους γειτονικούς μήνες. Στον επόμενο πίνακα εντοπίζεται πολύ μεγαλύτερη ομοιομορφία στις παραδόσεις των allowances με το Μαρτη να δέχεται ένα ποσοστό.



(α′) % Συναλλαγών για Κατανομή Αδειών

(β′) % Συναλλαγών για Παράδοση Αδειών

Σχήμα 1.2: Ομοφιλία βαθμών

1.3.3 Κεντρικότητα

Η γνώση που έχουμε απόκτηση για τις ιδιοτροπίες του ETS παρέχει εικόνα που διευκολύνει την ταυτοποίηση των σημαντικών χαρακτηριστικών του δικτύου. Οι κυβερνητικές οντότητες είναι υπεύθυνες για την κατανομή και συλλογή αδειών, οι regulated υποχρεούνται να υποβάλουν άδειες στο τέλος της περιόδου, έχοντας την ελευθερία να συναλλάσσονται όπως επιθυμούν στο ενδιάμεσο και οι οικονομικές οντότητες εξ αποκλεισμού αποτελούν διαφορετική κατηγορία και επιθυμούν την απόκτηση κέρδους εντός του συστήματος. Στους περισσότερους μήνες οι κυβερνητικές οντότητες δεν αλληλεπιδρούν με τις υπόλοιπες στο σύστημα, ενώ στους μήνες από Δεκέμβρη ως Απρίλη εξυπηρετούν μεγάλο και κεντρικό ρόλο.

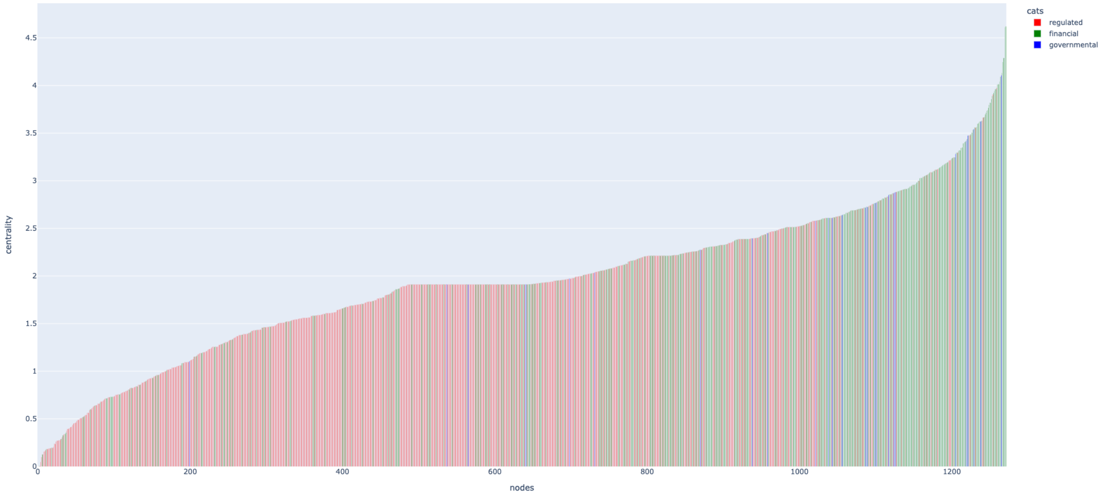
Για τον υπολογισμό σημασίας ενός κόμβου στο γράφο αναφέρουμε την ενδιάμεση κεντρικότητα όπως ορίζεται από τον Brandes [8]:

$$c_b(v) = \sum_{s,t \in V} \frac{\sigma(s,t|v)}{\sigma(s,t)}$$

με $\sigma(s,t)$ ο αριθμός των συντομότερων (s,t) μονοπατιών και $\sigma(s,t|v)$ ο αριθμός των συντομότερων (s,t) μονοπατιών που περνούν από κάποιο κόμβο v διαφορετικό των s,t. Αν $s=t$, τότε $\sigma(s,t)$, και αν $v \in \{s,t\}$ τότε $\sigma(s,t|v)=0$. Η μετρική ερμηνεύεται ως ο βαθμός στον οποίο ο κόμβος έχει έλεγχο επί των συνδέσεων του με άλλους κόμβους, βασισμένο στην υπόθεση ότι η σημασία των συνδέσεων είναι ίσα κατανεμημένη σε όλα τα συντομότερα μονοπάτια για κάθε ζεύγος.

Στο επόμενο γράφημα, απομονώνουμε το άνω 25% των κεντρικοτήτων του έτους 2014, περιορισμένο στις συναλλαγές 10-0,3-0,3-21. Για τη μικρότερη κεντρικότητα c_{min} του επιλεγμένου τμήματος: $\theta * c_{min} = 1 \rightarrow \theta = \frac{1}{c_{min}}$ Για κάθε κεντρικότητα c του διαστήματος αυτού $c' = c * \theta$

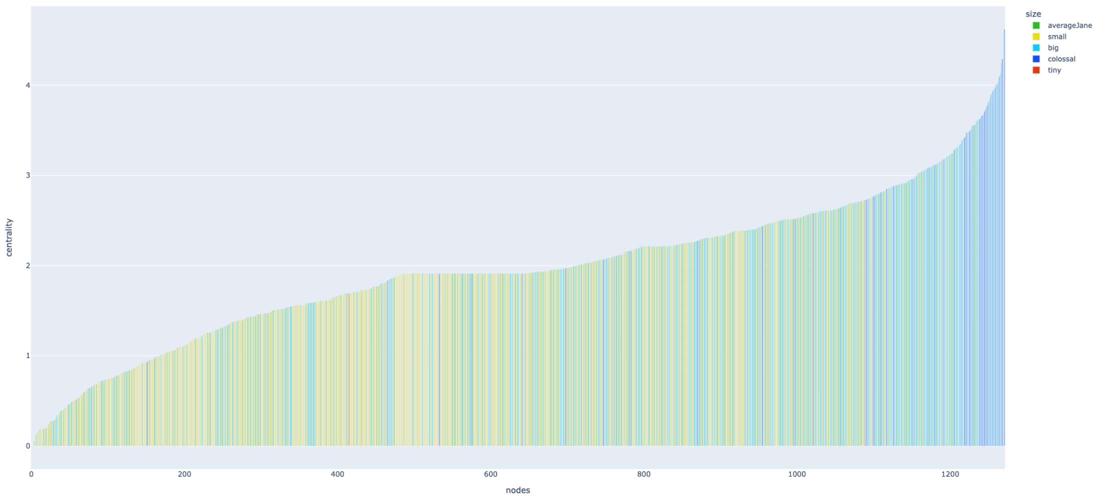
Λογαριθμώντας παίρνουμε το ακόλουθο γράφημα στο οποίο τα χρώματα στο υπόμνημα υποδεικνύουν την κατηγορία του κόμβο υ.



Σχήμα 1.3: Κεντρικότητα ως προς κατηγορία 2014

Φαίνεται πώς όταν δεν συμπεριλαμβάνουμε παραδόσεις και κατανομές αδειών, οι οικονομικής φύσεως κόμβοι και κάποιο κυβερνητικοί, που περιλαμβάνουν τα ιδρύματα πλειστηριασμών, είναι ο πυρήνας του δικτύου καθώς οι regulated κατέχουν θέσεις πιο κοντά στην περιφέρεια.

Στο κεφάλαιο 2 εισηγάγαμε την ορολογία ¨μέγεθος' βασιζόμενοι στο συνολικό αριθμό αδειών που συναλλάχθηκαν για κάθε οντότητα. Στην αναζήτηση επιπλέον πληροφορίας για την τοποθέτηση εντός του δικτύου θα μελετήσουμε το ιδιο γράφημα ως προς το μέγεθος.



Σχήμα 1.4: Κεντρικότητα ως προς μέγεθος 2014

1.3.4 Νόμος Δύναμης

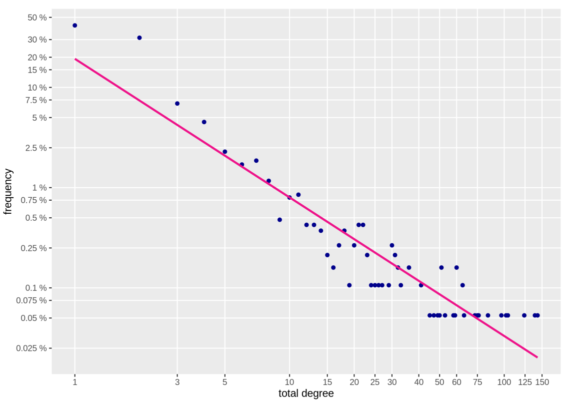
Σε πολλές επιστημονικές περιοχές, ένας συχνός ισχυρισμός που εμφανίζεται είναι πως τα δίκτυα ακολουθούν power-law κατανομές. Αυτό σημαίνει πως το ποσοστό των κόμβων p_k με βαθμό k ακολουθεί κατανομή $k^{-\alpha}$ με $a > 1$. Κάποιες εκδοχές έχουν πιο ισχυρές απαιτήσεις, όπως $2 < a < 3$ ή απόδειξη της εξέλιξεις με preferential attachment μηχανισμό.

Οι μέθοδοι για την εξέταση ύπαρξης power-law κατανομής πληθαίνουν και αυξάνονται σε ακρίβεια. Στη μελέτη τους το 2019 οι Broido et al. [9] κατασκευάζουν τεστ στο οποίο κατηγοριοποιούν σε 6 όρους, με βάση τα κριτήρια που πληρούν με αυτές να είναι, non scale-free , υπερ-αδύναμα, πολύ αδύναμα, αδύναμα, ισχυρά, πολύ ισχυρά. Με αυτές τις κατηγορίες εξετάουν 1457 δίκτυα, με 456 non scale-free , 431 υπερ-αδύναμα, 268 πολύ αδύναμα, 177 αδύναμα, 89 ισχυρά, 36 πολύ ισχυρά.

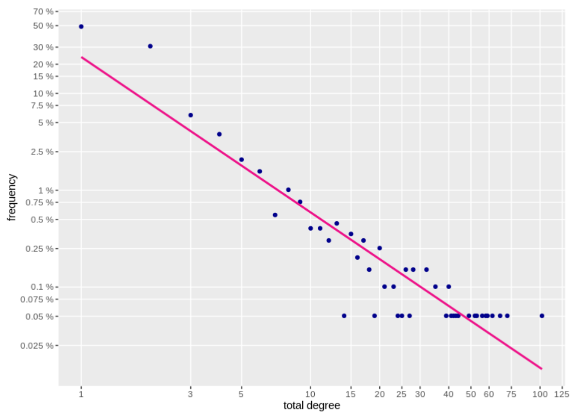
Ορίζοντας τον νόμο δύναμης, ο αριθμός των ακμών που έχει κάθε κόμβος, ή βαθμός κόμβου, χαρακτηρίζεται από την συνάρτηση κατανομής $P(k)$, η οποία δίνει πιθανότητα ένας τυχαία επιλεγμένος κόμβος να έχει ακριβώς κ ακμές. Η γενική μορφή του νόμος δύναμης είναι $P(k) \propto k^{-\gamma}$. Λογαριθμώντας παίρνουμε

$$P(k) = ak^{-\gamma} \Rightarrow \log P(k) = \log a - \gamma \log k$$

Δεδομένων των $P(k)$ και k από το δίκτυο, μέσω γραμμικής παλινδρόμησης με παραμέτρους $\log(a)$ και γ τη σταθερά και κλίση αντίστοιχα. Τα αποτελέσματα για p-value, R^2 και σφάλμα βρίσκονται στο παράρτημα. Ελέγχουμε το δίκτυο για κάθε μήνα για όλες τις συναλλαγές και για τις μειωμένες(3-0, 3-21, 10-0) και τα αποτελέσματα φαίνονται στα επόμενα σχήματα.

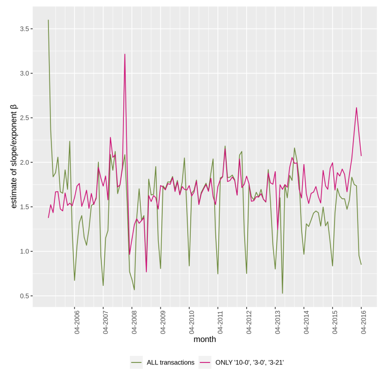


(α΄) Δεκέμβριος 2011

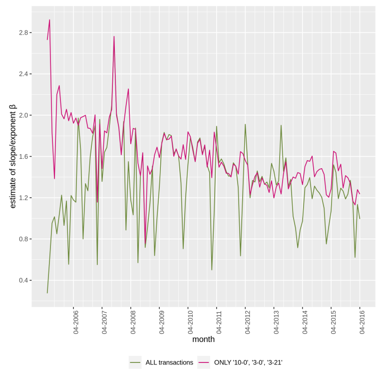


(β΄) Απρίλιος 2012

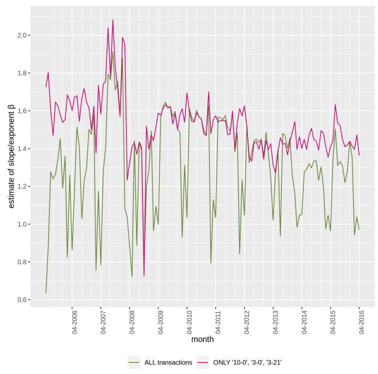
Σχήμα 1.5: Κατανομή Βαθμών



(α΄) Εκθετικό έσω-βαθμού

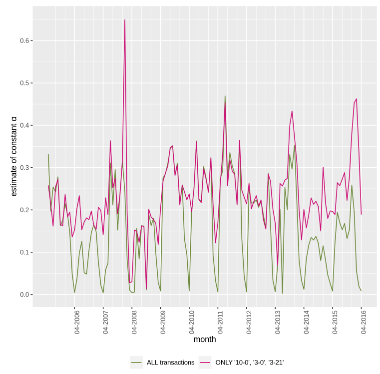


(β΄) Εκθετικό έξω-βαθμού

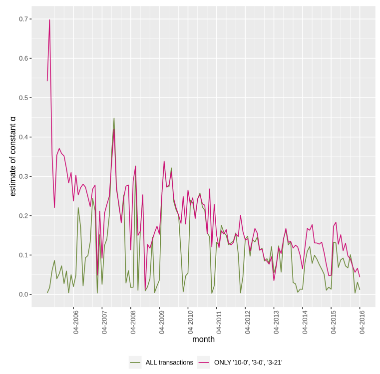


(γ΄) Εκθετικό συνολικού βαθμού

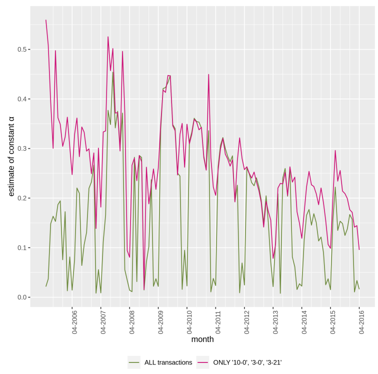
Σχήμα 1.6: Εκθετικά γραμμικής παλινδρόμησης



(α΄) Σταθερά έσω-βαθμού



(β΄) Σταθερά έξω-βαθμού



(γ΄) Σταθερά συνολικού βαθμού

Σχήμα 1.7: Σταθερά γραμμικής παλινδρόμησης

1.3.5 Μείωση Διαμέτρου

Όσο τα δίκτυα μεγαλώνουν με το χρόνο, η διαισθητική αναζήτηση τείνει στην εύρεση τρόπου αύξησης διαμέτρου είτε αυτό γίνεται με λογαριθμικό ρυθμό είτε με υπολογαριθμικό είτε με κάποιο ακόμα αργότερο ρυθμό. Οι Leskovec et al[37] στην εξέταση αυτής της διαίσθησης, απέδειξαν ότι στα δίκτυα στη δημοσίευσή τους στην πραγματικότητα έχουν μείωση της διαμέτρου.

Στη συνέχεια αυτής της υπόθεσης αναφέρουμε κάποιους αναγκαίους ορισμούς. Δύο κόμβοι στο δίκτυο είναι συνδεδεμένοι αν υπάρχει μονοπάτι μεταξύ τους. Για κάθε φυσικό αριθμό d, $g(d)$ το ποσοστό των συνδεδεμένων ζευγών κόμβων των οποίων το συντομότερο μονοπάτι έχει μήκος το πολύ d. Ο γραφος G έχει διάμετρο d αν το μέγιστο μήκος του συντομότερου μονοπατιού χωρίς ακμές για όλα τα συνδεδεμένα ζεύγη κόμβων είναι d.

Ουσιαστική Διάμετρος Για κάθε φυσικό αριθμό d, έστω $g(d)$ το ποσοστό συνδεδεμένων ζευγών κόμβων των οποίων το μέγιστο μήκος του συντομότερου μονοπατιού χωρίς ακμές για όλα τα συνδεδεμένα ζεύγη κόμβων είναι d. Έστω D ακέραιος για τον οποίο $g(D-1) < 0.9$ και $g(D) \geq 0.9$. Τότε ο γράφος G έχει ακέραια ουσιαστική διάμετρο D.

Δηλαδή, η ακέραια ουσιαστική διάμετρος είναι ο μικρότερος αριθμός βημάτων D στον οποίο τουλάχιστον 90% όλων των συνδεδεμένων ζευγών μπορεί να φταστεί.

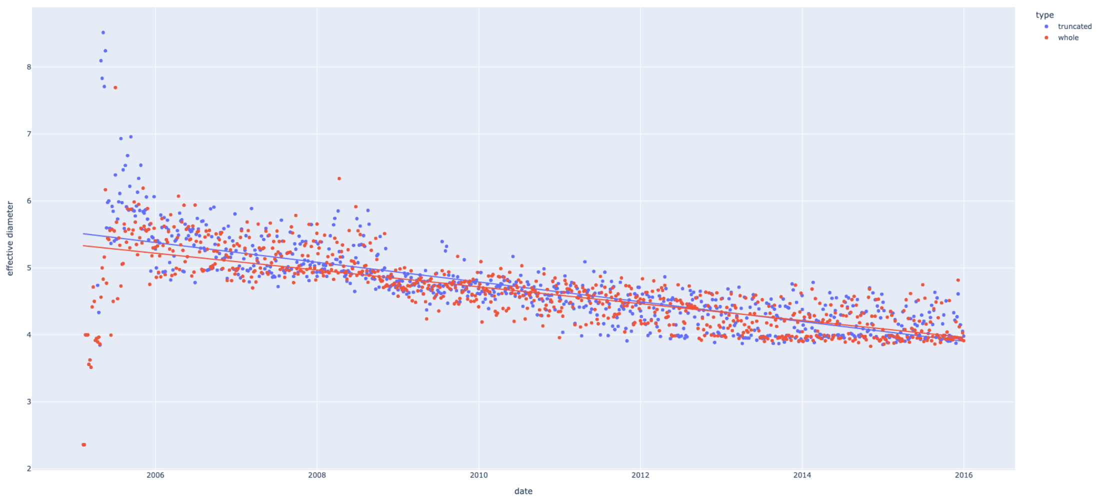
Επεκτείνοντας την παραπάνω συνάρτηση για να συμπεριλαμβάνει πραγματικές τιμές του x για $g(x)$ με γραμμική παρεμβολή μεταξύ των τιμών $g(d)$ και $g(d+1)$ $(d \leq x < d+1)$ παίρνουμε:

$$g(x) = g(d) + (g(d+1) - g(d))(x - d)$$

Εναλλακτικός ορισμός: Έστω D η τιμή όπου $g(D) = 0.9$ τότε λέμε ότι γράφος G έχει ουσιαστική διάμετρο. Ο ορισμός αυτός διαφέρει ελαφρώς από προηγούμενους ορισμούς στους οποίους ήταν η ελάχιστη ακέραια τιμή d έτσι ώστε τουλάχιστον 90% των συνδεδεμένων κόμβων έχουν απόσταση το πολύ d. Αυτή η παραλλαγή εξομαλύνει τον ορισμό επιτρέποντας μη ακέραιες τιμές.

Η χρήση της ουσιαστικής διαμέτρου προέρχεται από τη δύναμη της σε σύγκριση με τη διάμετρο(η μέγιστη απόσταση μεταξύ όλων των συνδεδεμένων ζευγών κόμβων) καθώς η διάμετρος υπόκειται σε εκφυλιστικά φαινόμενα όπως πολύ μεγάλες αλυσίδες κόμβων.

Εξετάζοντας το δίκτυο στην πλήρη του μορφή, τα σημεία που φαίνονται στο παρακάτω γράφημα είναι σε διαστήματα 5 ημερών, με τη γραμμή να αποτελεί τη γραμμή παλινδρόμησης ελαχίστων τετραγώνων. Με το μπλε υποδεικνύουμε την επιλογή των συναλλαγών 3-0,3-21,10-0 και αναφέρουμε ότι οι κλίσεις των καμπύλων είναι -0.0017 για το πλήρες και -0.002 για τις περιορισμένες συναλλαγές.



Σχήμα 1.8: Ουσιαστική Διάμετρος

Η ανάλυση στο σύνολο της διπλωματικής γίνεται κυρίως στο πλαίσιο στατικών μηνιαίων δικτύων. Για αυτό το λόγο εξετάζουμε αν ακολουθείται η ίδια συμπεριφορά σε μικρότερες χρονικές περιόδους χρόνου με σημαντικά λιγότερους κόμβους. Εξετάσαμε κάθε μήνα των ετών 2005-2015(ο Ιανουάριος του 2005 δεν είχε συναλλαγές) κρατώντας τις εμπορικές συναλλαγές(3-0,3-21,10-0) για την εξέλιξη της ουσιαστικής διαμέτρου. Τρέχοντας γραμμική παλινδρόμηση στις τιμές αυτών των μηνών φτάνουμε στα ακόλουθα αποτελέσματα. Σε 61 από αυτούς τους μήνες παρατηρείται αυξανόμενη ουσιαστική διάμετρος και στους υπόλοιπους 70 έχουμε μείωση, γεγονός αξιοσημείωτο δεδομένου πόσο μικρότερο πλαίσιο εξετάζουμε. Αυτό θεωρούμε ότι συμβαίνει καθώς οι νέοι κόμβοι έχουν την τάση να συνδέονται σε ήδη δημοφιλείς κόμβους, είτε κυβερνητικούς είτε οικονομικούς.

1.4 Μέτρα Σύγκρισης Δικτύων

1.4.1 Απόκλιση Πορτραίτων

Για τον ορισμό της Απόκλισης Πορτραίτων, χρειάζεται να οριστεί αρχικά ο B-πίνακας [4]. Η απόσταση μεταξύ δύο κόμβων u, v είναι ο μικρότερος αριθμός ακμών μεταξύ κόμβων και μπορεί να βρεθεί μέσω Breadth First Search(BFS) . Θεωρώντας κόμβο v_i, ένα $l-$κέλυφος είναι το σύνολο κόμβων $V_l \subseteq V$ κόμβων σε απόσταση l από τον v_i. Έτσι:

$B_{l,k}$ = ο αριθμός των κόμβων που έχουν ακριβώς k μέλη στα αντίστοιχα $l-$κελύφη τους

Από τον παραπάνω ορισμό μπορούμε να συμπεράνουμε ότι:

$B_{l,k} = NP_l(k)$

όπου N ο αριθμός των κόμβων και $P_l(k)$ το ποσοστό των κόμβων σε βαθμό της τάξης l.

Οι γραμμές του B μπορούν να ερμηνευθούν ως κατανομές πιθανότητας:

$$P(k|l) = \frac{1}{N} B_{l,k}$$

είναι η πιθανότητα ένας κόμβος που θα επιλέγει τυχαία θα έχει k κόμβους σε απόσταση l. Μια άμεση σύγκριση ανά γραμμή προκύπτει:

$$KL(P(k|l)||Q(k|l)) = \sum_k P(k|l)) = \sum_k P(k|l) log \frac{P(k|l)}{Q(k|l)}$$

όπου $KL(p||q)$ η Kullback-Liebler (KL) απόκλιση μεταξύ δύο κατανομών p, q, με Q ορισμένο για το δεύτερο πορτραίτο όμοια με $P(k|l)$ όπως ορίστηκε παραπάνω. Οι συγγραφείς δηλώνουν ότι υποφέρει από κάποια μειονεκτήματα όπως, μη ορισμένο για $KL(P(k|l)||Q(k|l))$ σε κάποιες περιπτώσεις και η έλλειψη συμμετρίας και επομένως δεν αποτελεί μέτρο απόστασης.

Σε προσπάθεια να διορθώσουν τα προβλήματα παρέχουν την KL απόκλιση όπως ορίζεται παρακάτω.

$$KL(P(k|l)||Q(k|l)) = \sum_{l=0}^{max(d,d')} \sum_{k=0}^{N} P(k,l) log \frac{P(k|l)}{Q(k|l)}$$

Έτσι η **Απόκλιση Πορτραίτων** είναι

$$D_{JS}(G, G') = \frac{1}{2} KL(P||M) + \frac{1}{2} KL(Q||M)$$

με $M = \frac{1}{2}(P + Q)$ μια μικτή κατανομή των P, Q ορισμένα από $P(k,l) = \frac{kB_{l,k}}{\sum_c n_c^2}$ όπως αναφέρθηκε παραπάνω.

Η Απόκλιση Πορτραίτων $0 \leq D_{JS} \leq 1$ παρέχει μία τιμή που ποσοτικοποιεί τη διαφορά των δύο δικτύων από τις κατανομές των αποστάσεων τους, με μικρότερη D_{JS} για πιο όμοια δίκτυα και μεγαλύτερη D_{JS} για λιγότερο όμοια δίκτυα.)

1.5 Μοντέλα Κατασκευής Δικτύων

1.5.1 Community Guided Agreement

Η προσέγγιση προκύπτει από την συνειδητοποίηση ότι οι power laws εμφανίζονται σε συνδυασμό με αυτό-όμοιες δομές, δηλαδή αντικείμενα που αποτελούνται από μικρότερα αντίγραφα του ευατόυ τους. Οι Leskovec et al [37]δείχνουν ότι ένα απλό ισορροπημένο δέντρο σταθερής διακλάδωσης b είναι αρκετό για να οδηγήσει σε power law πύκνωσης. Οι κόμβοι V του γράφου θα είναι τα φύλλα του δέντρου($n = |V|$ ανδ $n = b^H$). Ορίζουμε $h(v, w)$ ως την απόσταση μεταξύ 2 φύλλων v και w, δηλαδή το ύψος του κοινού τους προγόνου(το ύψος του μικρότερου υποδέντρου που περιέχει και τους 2 κόμβους).

Κατασκευάζουμε τυχαίο με τους κόμβους V ορίζοντας την πιθανότητα να προκύψει ακμή μεταξύ των v, w ως συνάρτηση f της ποσότητας $h(v, w)$. Αυτό θα αναφέρεται ως η συνάρτηση δυσκολίας και γίνεται εμφανές ότι θα πρέπει να μειώνεται με το h, αλλά υπάρχουν πολλές μορφές που θα μπορούσε μια τέτοια συνάρτηση να πάρει.

Η μορφή της f που αποφάσισαν ότι δουλεύει καλύτερα προέρχεται από τις υποθέσεις αυτό-ομοιότητας. Θα πρέπει να είναι scale-free , επομένως το $\frac{f(h)}{f(h-1)}$ θα πρέπει να είναι ανεξάρτητο του επιπέδου του δέντρου άρα και σταθερό. Αυτό συνεπάγεται ορισμό του τύπου $f(h) = f(0)c^{-h}$. Για λόγους απλότητας θέτουμε $f(0) = 1$. Έτσι έχουμε συνάρτηση δυσκολίας:

$$f(h) = c^{-h}$$

όπου $c \geq 1$ όπου c αναφέρεται ως σταθερά δυσκολίας.

1.5.2 Forest Fire

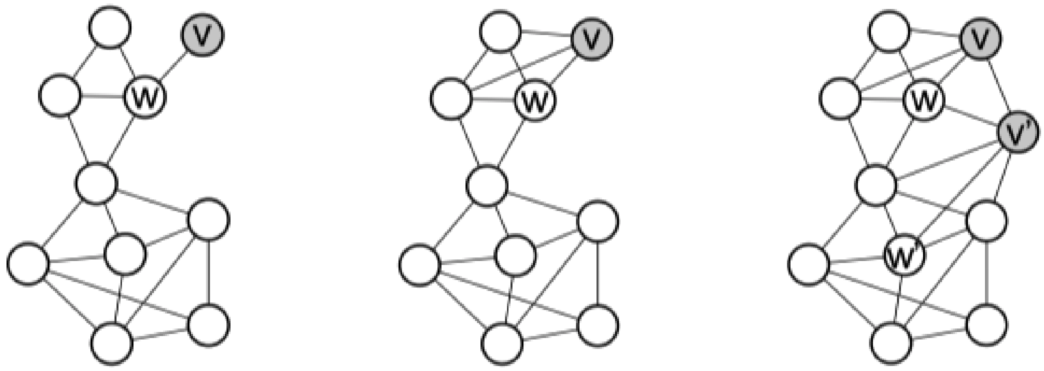
Για την αποτύπωση χαρακτηριστικών τα οποία το CGA δεν μπόρεσε, δημιούργησαν το μοντέλο Forest Fire .Οι κόμβοι καταφθάνουν σειριακά και σχηματίζουν έξω-ακμές με κάποιο υποσύνολο των υπάρχοντων κόμβων. Για να σχηματίσει ακμές, ο καινούργιος κόμβος v συνδέεται με κόμβο w στο υπάρχον γράφημα και ξεκινά να ¨καίει' ακμές από το w προς τα έξω, δημιουργώντας ακμή με κάθε καινούργιο κόμβο που ανακαλύπτει. Ο αλγόριθμος είναι ως εξής:

Ξεκινάμε με δύο παραμέτρους, p πιθανότητα προοδευτικής καύσης, r πιθανότητα αντίστροφης καύσης.

- ο v επιλέγει πρέσβη w τυχαία και σχηαμτίζει ακμή με τον w

- παράγουμε x από διωνυμική κατανομή με μέσο $(1-p)^{-1}$. Ο v επιλέγει x ακμές του w, επιλέγοντας έσω-ακμές με πιθανότητα r φορές λιγότερο από έξω-ακμές. Έστω $w_1, w_2, ..., w_x$ τα άλλα άκρα των επιλεγμένων ακμών

- ο v σχηματίζει έξω-ακμές με $w_1, w_2, ..., w_x$ και εφαρμόζει το (ii) αναδρομικά στους $w_1, w_2, ..., w_x$ αποφεύγοντας επαναλήψεις κόμβων



Σχήμα 1.9: Η διαδικασία του Forest Fire
Αριστερά: καινούργιος κόμβος v εισέρχεται στο δίκτυο, επιλέγει κόμβο w. Μέση: Ο v ενσωματώνεται συνδεόμενος αναδρομικά με τους γείτονες του w, τους γείτονες των γειτόνων, κ.ο.κ. Δεξιά: Ο v' αν δημιουργήσει μεγάλη φωτιά, συνδέεται με μεγάλο αριθμό των υπαρχόντων κόμβων

1.5.3 Preferential Attachment

Πολλά κοινωνικά δίκτυα χαρακτηρίζονται από άνισες κατανομές ακμών. Οι παρατηρούμενες ασύμμετρες κατανομές έχουν σε πολλές περιπτώσεις αποδοθεί σε preferential attachment , μία τάση των κόμβων σε ένα επεκτεινόμενο δίκτυο να σχηματίσουν νέες ακμές με προτίμηση σε κόμβους με υψηλό αριθμό ακμών. Όπως έχει διατυπωθεί πολλές φορές προηγουμένως στη βιβλιογραφία, οι μηχανισμοί preferential attachment δημιοργούν κατανομές που προσεγγίζουν power-law . Όπως έχει αποδειχθεί σε προηγούμενο κεφάλαιο, το δίκτυο μας ακολουθεί power-law κατανομή βαθμών, πράγμα που μας οδηγεί στην εξερεύνηση μοντέλων που αναπαράγουν τη δομή του δικτύου.

Μοντέλο του Price

Το 1965, ο φυσικός και ιστορικός της επιστήμης Derek de Solla Price περιέργραψε πιθανώς το πρώτο παράδειγμα αυτού που αποκαλείται scale-free δίκτυο[47]. Μελετώντας δίκτυα αναφορών σε επιστημονικά papers , παρατήρησε ότι οι έσω-βαθμοί και έξω-βαθμοί ακολουθούν power-law κατανομές. Ο Price ονόμασε αυτό το φαινόμενο cumulative advantage , με αυτό να γίνεται ευρύτερα γνωστό ως preferential attachment όταν μελετήθηκε αργότερα από τους Barabasi, Albert . Το μοντέλο που πρότεινε ο Price είναι το εξής:

Θεωρούμε κατευθυνόμενο γράφο με n κορυφές και p_k το ποσοστό των κόμβων με έσω-βαθμό k τέτοιο ώστε $\sum_k p_k = 1$. Νέες κορυφές φτάνουν συνεχώς στο δίκτυο και σε κάθε κόμβο που φτάνει ανατίθεται ένας αριθμός για τον έξω-βαθμό στη δημιουργία της κορυφές. Ο έξω-βαθμός ποικίλει μεταξύ κόμβων και ο μέσος έξω-βαθμός είναι m είναι σταθερός. Ο μέσος έσω-βαθμός είναι επίσης m καθώς $\sum_k kp_k = m$.

Η πιθανότητα ένας νεοαφιχθής κόμβος, π.χ. ένα νέο paper κάνει αναφορά σε υπάρχον συνδέεται σε υπάρχουσες κορφυές, είναι ανάλογη με τον έσω-βαθμό k παλιάς κορυφής v_i. Καθώς κάθε κόμβος ξεκινάει με 0 έσω-βαθμό, θα έχει διαρκώς μηδενική πιθανότητα να αποκτήσει καινούργιες ακμές. Για να παρακαμφθεί το πρόβλημα η πιθανότητα θα πρέπει να είναι αναλογική στο $k + k_0$. Μια σύμβαση, η οποία μπορεί να δικαιολογηθεί θεωρώντας την αρχική δημοσίευση ενός paper τον εαυτό του, θέτοντας $k_0 = 1$. Έτσι η πιθανότητα μία νέα ακμή να συνδεθεί σε έναν από τους υπάρχοντες κόμβους με βαθμό k είναι:

$$\frac{(k+1)p_k}{\sum_k (k+1)p_k} = \frac{(k+1)p_k}{m+1}$$

Όπως δείχνει ο Newman [43]:

$$p[k] \approx k^{-(2+1/m)}$$

Για μεγάλα n, η κατανομή βαθμών έχει ουρά power-law με εκθετικό $\alpha = 2 + \frac{1}{m}$. Αυτό συνήθως παράγει εκθετικά στο $(2, 3)$. Η παραπάνω εξίσωση όπως φαίνεται δεν εξαρτάται από το k_0 και έτσι η παράμετρος $k_0 = 1$ μπορεί να δικαιολογηθεί εκ των υστέρων.

Barabasi-Albert

Η δουλειά του Price παρέμεινε σε σχετική αφάνεια στηνε επιστημονική κοινότητα και το cumulative advantage δεν εγκαθιδρύθηκε σαν έννοια μέχρι που οι Barabasi και Albert [5] του έδωσαν καινούργιο όνομα, preferential attachment . Η διαφορά μεταξύ των δύο μοντέλων είναι η έλλειψη κατεύθυνσης στο μοντέλο των Barabasi, Albert οπότε δεν υπάρχει διάκριση μεταξύ έσω και έξω-βαθμών. Αυτό έρχεται με το μειονέκτημα της έλλειψης κατεύθυνσης ακμών, αλλά παρακάμπτουμε το πρόβλημα της αρχικής αναφοράς.

Η πιθανότητα μία καινούργια ακμή να συνδεθεί με κόμβο βαθμού k το ισοδύναμο της εξίσωσης του Price είναι

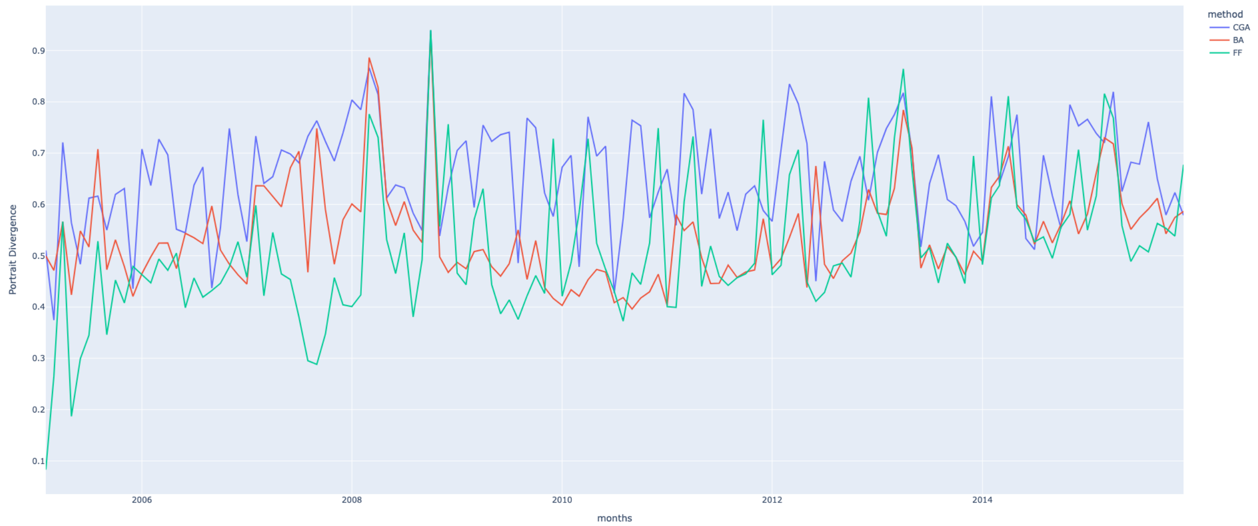
$$\frac{kp_k}{\sum_k kp_k} = \frac{kp_k}{2m}$$

Όπως αναφέρεται από τον Barabasi στο βιβλίο του το 2013 [6] υπάρχει ένας αριθμός αναλυτικών εργαλείων για τον υπολογισμό της κατανομής βαθμών του μοντέλου. Χρησιμοποιώντας continuum theory προβλέπεται:

$$p(k) \approx 2m^{1/\beta}k^{-\gamma}$$

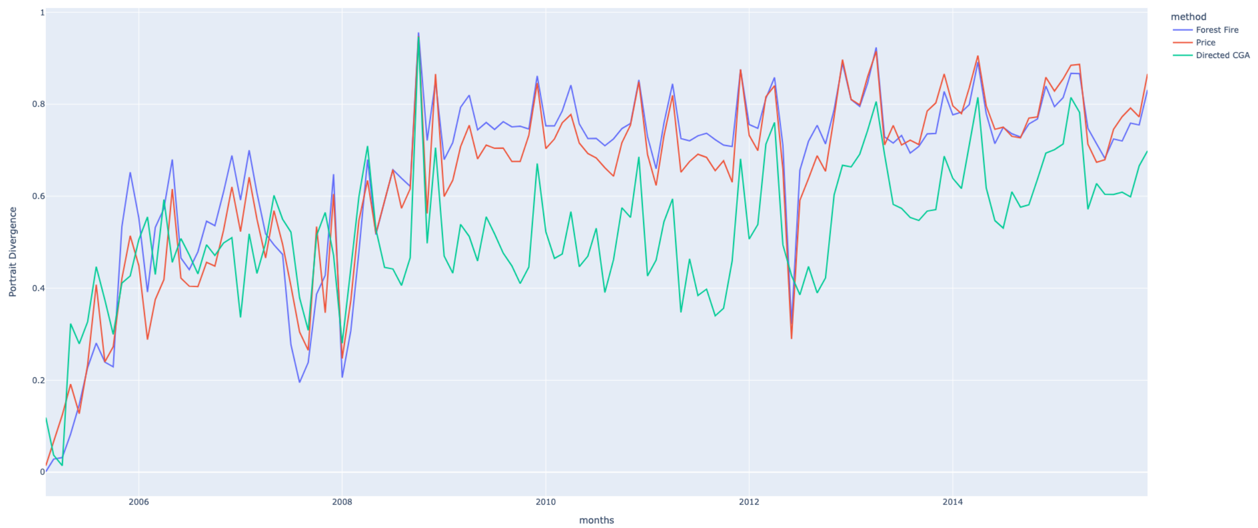
1.5.4 Αποτελέσματα Συγκρίσεων

Undirected μοντέλα



Σχήμα 1.10: Σύγκριση Undirected μοντέλων

Directed μοντέλα



Σχήμα 1.11: Σύγκριση Directed μοντέλων

1.6 Συμπεράσματα

Συνοψίζοντας, δείξαμε ότι ενώ στα περισσότερα δίκτυα υπάρχει ομοφιλία βαθμών, στο ETS οι κόμβοι μεγάλων βαθμών τείνουν να συναλλάσσονται με κόμβους μικρότερων βαθμών. Όταν εξετάσαμε τη συμπεριφορά των οντοτήτων με βάση τα γνωρίσματα τους, προκύπτουν patterns στα οποία παρουσιάζεται τάση να συναλλάσσονται με κόμβους της ίδιας χώρας και τις περισσότερες φορές με διαφορετικής κατηγορίας κόμβους.

Τα δίκτυα παρουσιάζουν ισχυρά στοιχεία ετήσιας περιοδικότητας και προβλεψιμότητας όσον αφορά τη δομή και την ποσότητα συναλλαγών, παρά τις αλλαγές στις φάσεις. Ακολουθώντας την τάση της βιβλιογραφίας, εξετάσαμε αν το δίκτυο ακολουθεί power-law στην κατανομή βαθμών και σχεδόν χωρίς εξαίρεση υπακούει την προαναφερθείσα κατανομή, αν και με μικρό εκθετικό.

Μια σημαντική συνειδητοποίηση όταν μελετάμε δίκτυα είναι η εξέλιξη τους στο χρόνο. Στη μελέτη της προόδου του δικτύου στο χρόνο αποδείχθηκε ότι αυξάνεται σε πυκνότητα καθώς και ότι μειώνεται σε (ουσιαστική) διάμετρο.

Στη συνέχεια εξετάσαμε κάποια υποψήφια μοντέλα για την αναπαραγωγή του δικτύου τα οποία συγκρίναμε και αξιολογήσαμε την απόδοση τους με βάση το μέτρο σύγκρισης Απόκλιση Πορτραίτων. Στην περίπτωση των undirected εκδοχή, το μοντέλο Barabasi Albert το οποίο τροποποιήσαμε για την περίπτωση μας. Το συγκρίναμε με το Community Guided Agreement και το undirected Forest Fire , με το BA να αναπαράγει τη δομή του δικτύου καλύτερα στις πιο σταθερές φάσεις. Στην κατευθυνόμενη περίπτωση, το μοντέλο του Price συγκρίνεται με το Forest Fire και με το κατευθυνόμενο CGA με το τελευταίο να επιτυγχάνει πολύ καλύτερες τιμές Απόκλισης Πορτραίτων

Ενώ η δουλειά μας δεν είναι η πρώτη που εξετάζει το EU ETS, παρέχουμε την πιο ολοκληρωμένη εικόνα για τα δικτυακά χαρακτηριστικά του. Παρά τη σφαιρικότητα της διπλωματικής, δεν αποτελεί πλήρη ανάλυση όλων των γνωρισμάτων του δικτύου και επιδέχεται επέκταση, με το κομμάτι της μοντελοποίησης να είναι ανοικτό στην εφαρμογή περαιτέρω μοντέλων. Ο τρόπος επεξεργασίας του δικτύου αποτελούταν από στατικές εικόνες ανά ορισμένα χρονικά διαστήματα. Μια ενδιαφέρουσα προσέγγιση είναι η προσέγγιση και μοντελοποίηση μέσω χρονικής εξέλιξης του δικτύου, με τρόπους παρόμοιους με εκείνους που αναφέρει ο Leifert [?] που εξετάζει τα TERGM(Temporal Exponential Random Graphs), SAOM(Stochastic ActorOriented Model) . Με τη γνώση και κατανόηση που έχει παρουσιαστεί σε αυτή την εργασία, ένα επόμενο βήμα θα μπορούσε να είναι η εφαρμογή μεθόδων Θεωρίας Παιγνίων για προσέγγιση της συμπεριφοράς των οντοτήτων του δικτύου.

Chapter 2

Introduction

2.1 Background

For the last half century the world has been in search of solutions for the problem of climate change, either managing or ameliorating its effects on the planet. Having widely recognized that the main cause of it are Greenhouse Gases (GHG) many of the efforts focus on limiting emissions levels. The world's response to the problem has been the creation of the Intergovernmental Panel on Climate Change (IPCC) in 1988 which was endorsed by the United Nations. The evolution of the aforementioned intergovernmental body has been the United Nations Framework Convention on Climate Change which was established in 1992 and was joined by 154 nations. Its signatories' commitment to reduce the atmospheric concentration of greenhouse gases were driven by the goal to prevent dangerous anthropogenic interference with Earth's climate system. The famous "20-20-20", is a legislative package of energy and climate targets set by the European Union to ensure that European Countries will able to reduce GHG emissions at least 20% below 1990 levels, to increase energy consumption from renewable sources by 20% and to improve energy efficiency by 20% until 2020.

The EU, in an effort to curb the emission problem, chose a "cap and trade" system as the best means of meeting the GHG emissions reduction target at the least overall cost to participants and the economy as a whole. Cap-and-trade was first theorized in studies between 1967-1970 [10] [11] as an approach to air pollution abatement.

The European Emissions Trading System caps the total volume of GHG emissions of all participants in the system. The ETS legislation creates allowances which are essentially rights to emit GHG emissions equivalent to the global warming potential of 1 tonne of CO_2 equivalent(tCO_2e). For every period set by the European Union, the participants are legally obligated to surrender allowances respective to their emissions.

The choice of this type of system was founded on the benefits that the structure exhibits. The certainty about the maximum quantity of GHG emissions for the period of time over which the system is set is reassuring the creators of the system that the international objectives and environmental goals will be achieved. The flexibility that trading brings means that all firms face the same carbon price and it ensures that emissions are cut where it costs least to do so. These types of systems provide the added benefit of generating an income for the regulator if GHG emissions allowances are auctioned and that can be invested in climate-change-oriented investments for the EU or for the member states. Finally the EU ETS provides certainty to emissions reduction from installations responsible for around 50% of EU emissions. This reduces the risk that Member States will need to purchase additional international units to meet their international commitments under the Kyoto Protocol [50].

2.2 Motivation

From the early 20th century scientists and engineers have been systematically trying to understand and break down networks for several types; social, financial, computer, biological, etc. The understanding we gain is crucial to making any breakthrough and is essential for progress in the field. In this thesis we attempt to provide a well-rounded framework that can be applied to a wide range of networks and is a solid base upon which interpretation of network structure can be founded.

Apart from the importance of the results we will demonstrate, this thesis is setting the groundwork for further expansion in future projects and serves as an effort in gaining knowledge of the behavior of the EU ETS network as far as its structure is concerned, but more importantly about the way the entities that participate in it behave the way they do. The utility this provides is multifaceted but the principal benefit we aspire to gain, is the ability, given a certain degree of stability, to forecast as accurately as possible the structure of the ensuing months and years.

Applying this type of analysis will also serve as a precursor to expanding this work to incorporate a game theoretic component in which with the now more clearly defined concepts and understanding, the approach might be ripe.

2.3 Previous Work

The work of this thesis is following a long line of work on the subject of network analysis. The consensus suggests that Leontief's 1936 economic input-output analysis [36] and Shannon's 1948 information theory [48] as the origins of many network analysis techniques. Hafner-Burton et al. [23] have a seminal paper in the field that provides a holistic review of methods to apply to network analysis when examining approaches to international relations networks such as transnational advocacy networks(TANs) as well as terrorists and criminals organized in "dark" networks. In the field of biology Kay et al. [28] provide an early background on steady-state ecological systems using flow network representation to describe the function of the ecosystem.

As will be referred to again later, Newman's [42] paper on mixing patterns in networks has provided useful tools such as mixing matrices and discussion on assortativity in networks which appear frequently in this thesis as well. Jackson [25] in his very thorough 2003 survey, takes a look at the literature on network formation, providing definitions of network games and discusses some of what is known about the (in)compatibility of overall societal welfare with individual incentives to form and sever links. Similarly and more recently de Paula [15] gives a selective literature review on econometric models of network formation and presents a discussion on dyadic and non-dyadic models, also providing a great background on the formation process. In a paper of the same name Chandrasekhar [12] looks first at the economic processes operating on network such as social learning and labor market search and second at how and why economic networks form. Studying these objectives he focuses on the econometric issues that arise when an empirical researcher seeks to model formation and the ways the researcher can tackle those issues.

In their 2019 paper, Goldner et al. [22] study the design of carbon license auctions within the EU ETS. They use facts regarding uniform price auctions to understand what is happening in these markets and probabilistic analysis to handle the uncertainty of valuation realizations.They,then, apply a Price of Anarchy analysis to give strategic guarantees and provide concrete recommendation, with provable approximation guarantees, for how to set the parameters of the mechanism used in practice. Using a similar approach, García-Algarra et al. [21] examine its topological and statistical properties and with those results they provide generative models that closely mimic the properties of annual empirical data. In their 2011 paper, Akerman et al. [1] when examining the global arms trade network follow a very similar route to ours, by studying the network structure and trying to identify node importance, interactions, degree distribution and network modeling.

The EU ETS has been one of the largest carbon markets in operation. Its importance has led to extensive analysis in the literature. Performing a network based analysis of the European Emission Market, Karpf et al [26] note that the network exhibits a strong core-periphery structure also reflected in the network formation process. This thesis follows a similar trajectory, by examining what we deem as key characteristics and using the evidence gained to provide modeling of the network. However as will be clear by the following chapters we provide further insight and more complete creation and comparison. In a more thorough study on the EU ETS was released in 2018 Karpf et al. [27] examine various facets of the system. In particular they conduct a data-driven approach to the network in which they consider a static network up until the year 2011. The paper focuses on identifying and analyzing the characteristics of the network along with their influences on its structure, using an ERGM fitting to emphasize the importance of each attribute. They further propose a dynamic model to capture the changes over time. In the static network, the paper lacks in examination of models and in the direct comparison with the real network. It also suffers from data gathered until 2011 consisting of the first and part of the second period of the EU ETS which with hindsight from the 3rd period, the period they examine appears wildly inconsistent and contains periods whose allocation is based on outdated estimates of the need for allowances. In a similar vein, Borghesi et al. [7] analyzing the system from a country-level perspective instead of firm-level perspective, arrived at the finding that Person Holding Accounts- which approximate the intermediaries in the network- play a prominent role in the network using a variety of centrality measures like PageRank, in/out strength, average neighborhood in/out strength, degree centrality and others.

Our focus is mainly on the network's attributes and properties and give models that closely resemble the structure of the real network. However the EU ETS has been studied from other perspectives such as the auction prices and allowance prices estimates and equations. Dimos et al. in their 2020 paper [18] study the the effects of allowance banking and the financial sector in the EU ETS where they find that it is a considerable though not dominant price determinant. Extending their work on the EU ETS [17] provide a definition of a subset of nodes, named All Time Almost Dominating Set (ATADOM), the nodes of which emerge as the core transactors. Using a Vector AutoRegressive model they attempt to forecast the EUA price.

Another way to view the EU ETS is as a temporal network, that is a network evolving over time. This approach is not as common, however a framework has been laid out by Leifeld et al. [34] who compare the Temporal Exponential Random Graph Model with the Stochastic Actor Model that are similar in their mathematical core. In the same field, Holme [24] recognizes the shortage of tools and the relative youth of the field and provides a review of the methods to analyze and model temporal networks and processes taking place on them.

2.4 Contribution

The EU ETS has been, as will be expressed in length in the next chapter, analyzed many times from various perspectives. However, with the clause of not releasing transactions until three years to the date after they occur, there is a serious lack of network analysis studies on this system. Also, while there is an extensive literature on the price of allowances, the impact of the economic recession on the European power sector emissions [16], the driving forces of price [14] and many more price related papers, there are very few that provide models for the network structure . The most complete such study was performed by Karpf [27] in 2018. In this thesis we provide a thorough exploration of the characteristics of the network perspective of the system and as far as the literature is concerned we haven't encountered any such study being done on the EU ETS and few studies have covered as many areas. Our analysis outlines a methodology for understanding structural characteristics of large networks and relationships between different types of node classifications.

As far as the specific aspects of the EU ETS, we come to a clear picture on the preferences of each category of network nodes obtaining knowledge not previously expressed as detailed. An essential part of the structure is the correlation between the degree of each node and its proclivity to connect with other nodes based on their degrees. We reach a surprising conclusion not usually found in network analysis particularly in financial and social networks, as large degree nodes have the tendency to connect to other large degree nodes. On the contrary, our results depicted also in the images the images from monthly graphs indicate that the larger nodes form a community with nodes of smaller degrees. We manage to statistically prove the repetitiveness of patterns in months with statistical measurements which fall in line with the information we had about the system beforehand.

As is customary with network analysis studies and surveys we prove that the EU ETS follows a power-law degree distribution which helps the modelling of the network in later chapter. Following in the footsteps of Leskovec[37] who tried to model scale-free networks using not so straightforward rich-get-richer method of modelling networks, we prove that the effective diameter of the network decreases over time as well that there exists a parameter α such that the number of edges follows a power law to the number of nodes with α as exponent, proving the densification law as stated by Leskovec. We haven't observed the examination of any such hypotheses in other studies concerning the EU ETS.

While many papers and surveys arrive at a point of fitting a model to the network, we have seen few occasions where the network is created using the stochastic process and then compared using some comparative measure as most settle on the fitting. In this thesis, using Portrait Divergence[3] which is a very useful tool for comparing structural differences between networks, we show comparative results for models, which are the natural conclusion of the previous analysis, both for the directed and the undirected case of graphs. The results show that in the undirected case, the network can be modelled best, in the most stable phases, by the Barabasi-Albert model[5] when we alter it to include non integer values for the model parameter m while in the directed both models we test, the directed version of the Community Guided Agreement seems to outperform the Forest Fire[37] and Price's models[47].

2.5 Outline

The following chapters and main part of the thesis are structured as follows. In the second chapter there will be an overview of the EU ETS, consisting of a thorough description of its main parts, its history including a relative literature that either has been in the direction of this thesis, or relative to the system and contributes in some way. In chapter 4 we analyze essential concepts of the transaction network and the system in general, focusing on the results we deemed most significant. Paving the way for chapter 6, in chapter 5 there will be comparison algorithms through which will come the evaluation of the models. Finally in chapter 6, we have surveyed the literature and have chosen suitable models to approach the networks and exhibit the results that arise when comparing the said models with the reality of the EU ETS.

Chapter 3

EU ETS

3.1 Introduction to the ETS system

In the last half century, climate change has been a major issue plaguing humanity. The world's response to the problem has been the creation of Intergovernmental Panel on Climate Change(IPCC) in 1988 which has been endorsed by the United Nations. The evolution of previously mentioned intergovernmental body has been the United Nations Framework Convention on Climate Change which was established in 1992 and was joined by 154 nations. Its signatories' commitment to reduce the atmospheric concentration of greenhouse gases were driven by the goal to prevent dangerous anthropogenic interference with Earth's climate system. The European Emissions Trading System is a 'cap and trade' system, in that it caps the total volume of GHG emissions of all participants in the system. The ETS legislation creates allowances which are essentially rights to emit GHG emissions equivalent to the global warming potential of 1 tonne of $CO_2(tCO_2e)$. [51]

The choice of this type of system was founded on the benefits that the structure exhibits. The certainty about the maximum quantity of GHG emissions for the period of time over which the system is set is reassuring the creators of the system that the international objectives and environmental goals will be achieved. The flexibility that trading brings means that all firms face the same carbon price and ensures that emissions are cut where it costs least to do so. These types of systems provide the added benefit of generating an income for the regulator, if GHG emissions allowances are auctioned and that can be invested in climate-change-oriented investments for the EU or for the member states. Finally the EU made it so the ETS provide certainty to emissions reduction from installations responsible for around 50% of EU emissions. This reduces the risk that Member States will need to purchase additional international units to meet their international commitments under the Kyoto Protocol.

3.2 Other Cap and Trade Systems

3.2.1 Regional Greenhouse Gas Initiative

The Regional Greenhouse Gas Initiative(RGGI) is a consortium of northeastern US states Connecticut, Delaware, Maine, Maryland, Massachusetts, New Hampshire, New Jersey, New York, Rhode Island, Vermont and Virginia so far, that limit carbon dioxide emissions from electricity generation through a regional emissions trading program. It came into effect in mid-2008 and since then total emissions from the region's power sector have dropped substantially. Being almost as old as the EU ETS, there is extensive literature on it as Murray et al. [41] in their 2015 econometric study developed a model to simulate base emissions and provide concrete policy implications. Looking at different aspects of the system Yan [53] estimated the impact to the regulated and the non-regulated neighboring states while Kim et al. [29] examined the effects of the system on emissions and on fuel-to-gas switching.

3.2.2 California and Quebec

California and Quebec are part of cap-and-trade systems that are linked via the guidelines of the Western Climate Initiative(WCI), a voluntary subnational governmental organization initiated in 2007. With the same goal of emission control, their goals are different with California pledging to reduce its emissions by 2020 to 1990 levels, while Quebec committing to reducing emission 20% below 1990 levels. The California cap-and-trade system covers around 75% of the statewide greenhouse gas emissions and the California Air Resources Board (CARB) has set a new standard of pursuing a legal mandate to reduce statewide greenhouse gas emissions at least 40% below 1990 levels by 2030.

3.2.3 Tokyo Cap and Trade

Since announcing its Tokyo Climate Change Strategy in June 2007, Tokyo Metropolitan Government (TMG) has been examining ways to bolster the fight against global warming. Nishidia et al. [44] study the outcomes of the Tokyo system, evaluating whether the goals set at the beginning of system are being met. The paper assesses the effectiveness of TCTP based on extensive data obtained from 1300 facilities covered by the programme and surveys among facility owners from the first compliance phase. They find that the date indicate that TCTP has been working effectively to reduce energy consumption in participating facilities to meet the ambitious emission reduction goals, to introduce new technologies, and to raise awareness and drive behavioural changes for energy demand reduction.

3.2.4 New Zealand Emissions Trading Scheme

The NZ ETS was first legislated in the Climate Change Response (Emissions Trading) Amendment Act 2008 in September 2008. This system broke some new ground in ETS design as noted by Leining et al.(2020) [35] who examined three keys areas of innovation, broad sectoral coverage with some upstream points of obligation, the absence of a hard limit on system emissions, and a two-part cost containment mechanism. Being a recent paper they had the luxury of 10 years of data, they show that despite its problems such substantial technical and political challenges, it can serve as a source of insights that can inform the future development of emissions trading globally.

3.3 Brief History

In March 2000 the European Commission presented a green paper on "Greenhouse gas emissions trading with the European Union" providing some first ideas on the designs of the EU ETS. That served as the basis for numerous stakeholder discussions that helped shape the EU ETS in the first phases. That led to the adoption of the EU ETS Directive in 2003 and the introduction of the EU ETS in 2005.

The first phase of the EU ETS spanned from 2005 to 2007 and was used to test price formation in the carbon market and to establish the necessary infrastructure for monitoring, reporting and verification of emissions. As there was no emission data available the cap that was set in the first period was largely based on estimates. The primary purpose of the first phase was to serve as a pilot to ensure that the system would function effectively ahead of 2008, to ensure that it would allow the EU Member States to meet their commitments under the Kyoto Protocol. The Directive 2004/101/EC of the European Parliament and the Council of 27 October 2004[13] allowed businesses to use certain emission reduction units generated under the Kyoto Protocol mechanisms Clean Development Mechanism(CDM) and Joint Implementation(JI) to meet the obligations under the EU ETS. During the first phase businesses were only allowed to use units generated under the CDM for EU ETS compliance.

The second phase of the system ran from 2008 to 2012, the same period as the first commitment period under the Kyoto Protocol. From the start of the second phase businesses were given the ability to use emission reduction units generated under JI to fulfil their obligations under the EU ETS which made the EU ETS the largest source of demand for CDM and JI emission reduction units. Towards the end of phase 2 the scope of the EU ETS was expanded by including aviation from 2012.

The first two phases shaped the formatting of the third phase of the EU ETS. The third phase runs from 2013 to 2020 which coincides with the Kyoto Protocol second commitment period, as agreed in Doha in December 2012. The EU ETS does not have an end date and continues beyond phase 3. From 2021 we have entered into the fourth phase of the EU ETS which is set to last up to 2030.

3.4 Environmental Goals

In 2008 the EU set a series of climate and energy targets to be met by 2020 in its pathway towards a low-carbon competitive economy, known as the "20-20-20" targets. These are:

- A reduction in EU greenhouse gas emissions of at least 20% below 1990 levels
- 20% of EU energy consumption to come from renewable resources
- A 20% reduction in primary energy use compared with projected levels, to be achieved by improving energy efficiency.

The EU ETS will play a key role in promoting decarbonisation in sectors such as the power sector. The EU ETS has a default emission reduction of 1.74% per year that applies beyond 2020, with a review set to take place before 2025. The overall GHG emission reduction target of 40% outlined in the proposed 2030 framework implies an overall reduction of EU ETS emissions by 43% relative to 2005, equivalent to a linear emission reduction in of 2.2% per year beyond 2020. The cap would need to be met through domestic emissions reductions within the EU. This will allow the EU ETS to continue its significant contribution in moving to a low carbon economy by 2050.

3.5 The EU ETS Mechanism

The EU ETS is a 'cap and trade' system that works by capping the GHG emissions for all participants in the system. The legislation of the EU ETS creates allowances, each one essentially translating to the right to emit GHG emissions equivalent to the global warming potential of 1 tonne of CO_2 equivalent(tCO_2e). It is evident that the level of the overall cap determines the number of allowances available in the system. For the first two phases of the system the cap remained but starting from the third phase and is decreasing annually from 2013 reducing the number of allowances available to businesses covered by the EU ETS by 1.74% per year.

Each year, a proportion of the allowances are handed out to some participants for free, while the rest are sold mostly through auctions. When the years ends the participants of the system must return a proportional amount of allowances to that of tonnes of CO_2e they emit during that year. When a participant doesn't meet the requirements for allowances it must either buy allowances from the market or acquire them through the auction houses. Otherwise it is imperative to reduce its emissions.

The value of the allowances is set by their demand because there is a limited or capped supply and are wanted most by those participants for whom the cost of making reductions is higher than other participants. As such, it allows the effort to be redistributed between participants so that emissions reductions take place in areas where it costs less.

In order to ensure compliance there are penalties and an enforcement structure. If companies fail to comply with their allowance surrender obligations there are significant fines set at €$100/tC0_2$ and rising with EU inflation from 2013. In addition, firms face an obligation to surrender the allowances owed. Thus, the cap (i.e. the environmental target) is maintained effectively.

In the following sections we provide some foundations for the analysis and understanding of the EU ETS mechanism as described in the ETS handbook [52].

3.6 Allocation of Allowances

A fundamental component of the mechanism is the allocation of allowances, which is done either by free allocation, or by via auctioning of allowances. 5% of allowances are set aside for new entrants to the system, which according to the EU ETS handbook are new installations receiving a new permit after 30 June 2011 or existing installations, with significant capacity extensions after 30 June 2011 (the significant extension must be a physical change and adjustments in allocation according to new entrant rules).

3.6.1 The evolution of allocation

In the first two phases (2005-2007 and 2008-2012) most allowances were allocated for free to the participants. The amount of allowances each installation received was calculated via the NAP (National Allocation Plans). That means that each member state would prepare and publish a document called NAP which consisted of the proposed number of allowances to be allocated for its installations over the duration of trading period. Then these would be processed by the Commission, who would approve or amend the total number of allowances to be allocated, based on criteria set in the annex of the 2003 EU ETS Directive.

For the third phase (2013-2020) some changes were implemented to the existing format. Here member states are still required to acquire an "allocation plan", known as National Implementation Measures (NIM) document which contains all of the detailed information about the allocations planned for each installation in the country. The member states are responsible for collecting data and the final allocation and the commission is responsible for approving or rejecting the NIMs or parts thereof, requiring amendments where necessary.

3.7 Auctioning

Another component of the ETS is auctioning which is a transparent allocation method that allows market participants to acquire allowances concerned at the market price. During the first two periods the member states were in charge of auctioning emission allowances. In the first(2005-2007), the member states were allowed to auction up to 5% of the emission allowances and during the second(2008-2012) up to 10%. Member States only exercised this right marginally and in phase 2 only 4% of allowances were actually auctioned.

From the third trading period (2013-2020) onwards is governed by the Auctioning Regulation[18] which specifies the timing, administration and other aspects of how auctioning should take place to ensure an open, transparent, harmonised and non-discriminatory process. Auctioning can take place on a common auction platform appointed pursuant to a procurement procedure conducted by those Member States. The joint procurement approach is taken by the European Commission and 25 participating Member States. Germany, Poland and the UK chose to opt-out from the joint procurement procedure and have their own auction platform. The maximum duration for each appointment of auction platform is 5 years.

The European Energy Exchange AG (EEX) is the transitional common auction platform for 25 Member States, and is also, separately, the opt-out common auction platform for Germany. The other auction platform is ICE Futures Europe (ICE), which is the opt-out auction platform for the UK. Poland has so far not listed an opt-out auction platform, so it temporarily uses the transitional common auction platform EEX. Norway, Liechtenstein and Iceland also use the transitional common auction platform.

Each bidder may apply for admission to bid at the auction platforms from anywhere in the EU and the EEA-EFTA. The auction platform must check each application to ensure bidders are eligible to participate under the rules laid down by the Auctioning Regulation and to prevent the auctions being used for criminal activity.

To ensure fair and orderly auctioning, there are two levels of supervision:

- Scrutinising and monitoring by the auction platform itself;
- Supervision by the competent national authority for financial markets of the Member State where an auction platform is located.

In addition, for horizontal supervision of all auctions on all auction platforms, an auction monitor will be appointed through a joint procurement procedure involving all the Member States and the Commission.

3.8 Data Specifications

In this section it is imperative to go into some depth on the types of transactions that take place in the network. As seen in Table 2.1 there are 41 types of transactions with the types varying by location, function, group, etc. Every type of transaction is comprised of two numbers, first of which is the main with the second connotating the secondary. With main type "10" we observe the transaction that are performed with the European Union or European Commission and their subsidiaries as part of the EU ETS process. Such transactions include, the allocation of allowances over various periods, which as can be seen in the table below can have different secondary type for different phases, the surrenders which the EU receives, auctions performed by the EU auction houses, etc. With main type "2" we get the transactions that involve the transformation of a unit to create ERU(Emission Reduction Units) which under Joint Implementation(JI) are reduction in the rate of GHG emissions. Type "3" covers any external transaction between two registries, a broad umbrella under which most transactions are classified. Types "4","5" and "6" refer to the internal transfers of units within the subsidiaries of a registry corresponding to cancellation, retirement and replacements of units. The second type refers to a subcategory or specficiation of the main type, with the most notable being "2" for allowance surrender, "36","37","51","52","53", the issue and allocation of allowances, "0" for no supplementary type. In the next table we lay out the types of transactions ordered by the volume of the network they contain with additional information such as the percentage of all transactions they represent.

Table 3.1: Number of transactions and total volumes for each transaction type.

type	volume	% volume	transactions	% trans.	name of the 2nd part of the type
10-0	52,717,720,379	36.539 %	411,006	46.768 %	0-No Supp
10-2	21,157,732,356	14.664 %	130,371	14.835 %	2-Allowance surrender
3-0	17,782,880,273	12.325 %	155,180	17.658 %	0-No Supp
1-0	16,543,694,755	11.466 %	30	0.003 %	0-No Supp
10-53	16,254,700,214	11.266 %	82,367	9.372 %	53-Allowance allocation
1-51	3,902,966,882	2.705 %	93	0.011 %	51-Allowance issue (2005-2007)
10-52	3,901,221,237	2.704 %	17	0.002 %	52-Allowance issue (2008-2012 onwards)
10-36	3,409,722,337	2.363 %	36,982	4.208 %	36-Allocation of general allowances
10-37	2,345,413,000	1.626 %	781	0.089 %	37-Auction delivery
10-34	1,749,540,826	1.213 %	8,325	0.947 %	34-General allowances - Banking
3-21	989,019,687	0.685 %	15,366	1.748 %	21-External transfer (2005-2007)
10-90	578,037,810	0.401 %	269	0.031 %	90-Deletion of allowances
5-0	573,766,311	0.398 %	12	0.001 %	0-No Supp
3-2	491,053,113	0.340 %	6,967	0.793 %	2-Allowance surrender
10-71	410,166,902	0.284 %	9,823	1.118 %	71-Exchange
10-72	406,742,478	0.282 %	9,765	1.111 %	72-Exchanged
1-31	300,000,000	0.208 %	2	0.000 %	31-Issuance of general allowances
10-35	293,375,433	0.203 %	2,109	0.240 %	35-Allocation of aviation allowances
10-41	199,077,383	0.138 %	18	0.002 %	41-Cancellation and replacement
4-0	81,626,128	0.057 %	4,204	0.478 %	0-No Supp
10-1	68,485,722	0.047 %	3,272	0.372 %	1-Allowance cancellation (2005-2007)
1-22	36,964,946	0.026 %	2	0.000 %	22-External transfer between art63a registries
10-92	19,493,569	0.014 %	130	0.015 %	92-Reversal of Allowance Surrender
2-0	14,052,192	0.010 %	90	0.010 %	0-No Supp
10-4	10,956,682	0.008 %	352	0.040 %	4-Surrender Kyoto Units from AOHA
10-86	9,409,597	0.007 %	651	0.074 %	86-Reverse of Excess Allocation
10-82	8,503,102	0.006 %	77	0.009 %	82-Reversal of surrender
3-75	6,963,838	0.005 %	25	0.003 %	75-AAU set aside
10-33	5,303,040	0.004 %	355	0.040 %	33-Aviation allowances - Banking
4-2	2,715,659	0.002 %	21	0.002 %	2-Allowance surrender
10-93	1,316,081	0.001 %	51	0.006 %	93-Correction
3-82	1,139,829	0.001 %	5	0.001 %	82-Reversal of surrender
10-171	1,065,324	0.001 %	3	0.000 %	171-Reversal of Exchange
10-24	1,011,231	0.001 %	4	0.000 %	24-Issuance - Internal transfer Art 63a
1-24	1,011,231	0.001 %	4	0.000 %	24-Issuance - Internal transfer Art 63a
10-190	877,233	0.001 %	4	0.000 %	190-Reversal of deletion
10-136	751,724	0.001 %	46	0.005 %	136-Allocation of general allowances
10-26	508,510	0.000 %	20	0.002 %	26-Conversion of art63a allowances
4-26	508,510	0.000 %	20	0.002 %	26-Conversion of art63a allowances
4-91	1,415	0.000 %	2	0.000 %	91-Cancellation against deletion
10-135	18	0.000 %	2	0.000 %	135-Allocation of aviation allowances

3.9 Category Classification

The entities in the system are account holders each of which may contain a number of subsidiaries of various types. The classification we make here is based on the prevailing nature of each account holder.

Regulated entities

Regulated entities are actors with obligation to surrender allowances for their GHG emissions annually. An account holder is regarded as a regulated entity (or simply "regulated") if it holds an Operator Holding Account or an Aircraft Operator Account (i.e., if it operates a stationary installation or an aircraft).

Governmental entities

Governmental entities are the European Commission and other European Union accounts along with the ministries and administration bureaus of the countries participating in the ETS. The main role of governmental entities is to supply the system with allowances (free or sold at auction) and to receive the surrendered allowances. An account holder will be classified as a governmental entity(or simply "governmental") if it cannot be classified as regulated and it holds an administrative account. Moreover, the owners of the EEX and ICE auction platforms are also classified as governmental entities, as their main role is to supply the system with allowances through auctions.

Financial entities

Purely financial entities participate in the ETS to serve their own interests, and mostly store allowances or trade them with regulated or other financial participants. An account holder is classified as a purely financial entity (or simply "financial") if it cannot be classified as regulated or governmental and holds an account type associated with allowance trading. Note that regulated entities may also hold accounts associated with allowance trading.

3.10 Size Classification

The entities that take part in the EU ETS process, vary in the type of category, the sector the belong to as well as in their activity in the system. The so called regulated actors differ wildly in emissions by unit of product. Some possess a huge number of installations while others contain only a few or just one. Their resulting emissions are bound to differ as much as their installations, with the larger account holder having to surrender large amounts of allowances. This variation in the entities translates to a variation in the transacting volumes, trade activity and free allocation of allowances. It then becomes useful to provide some categorization for the transacting size of the agents. We observe that there are few agents that are heavy emitters. The majority exhibit a quite negligible behavior, but the most active agents may be susceptible to exhibit marketpower. As such we consider "size" of an agent in terms of average of annual volume of transacted allowances. From that we split the list of agents to five different levels of size which are: tiny, small, average(or averageJane), big and colossal.

Classification: Let q be the volume of allowances, i the agent and t the year. We note as $q_{i,t}$ the annual quantity for agent i at year t. We note $\mathbb{E}[X]$ the mean value of X(or the expectation, if X is random variable). We note $q_{i,t}|i$ q as the number of allowances of agent i while t varies. As a first step towards classification, we compute for every agent, the natural logarithm of the average annual number of allowances as

$$\mathbb{E}[q_{i,t}|i]$$

Partitioning that quantity to five equal intervals, we arrive to the aforementioned levels of size. For example an agent's size is tiny, if the quantity is within the first interval.

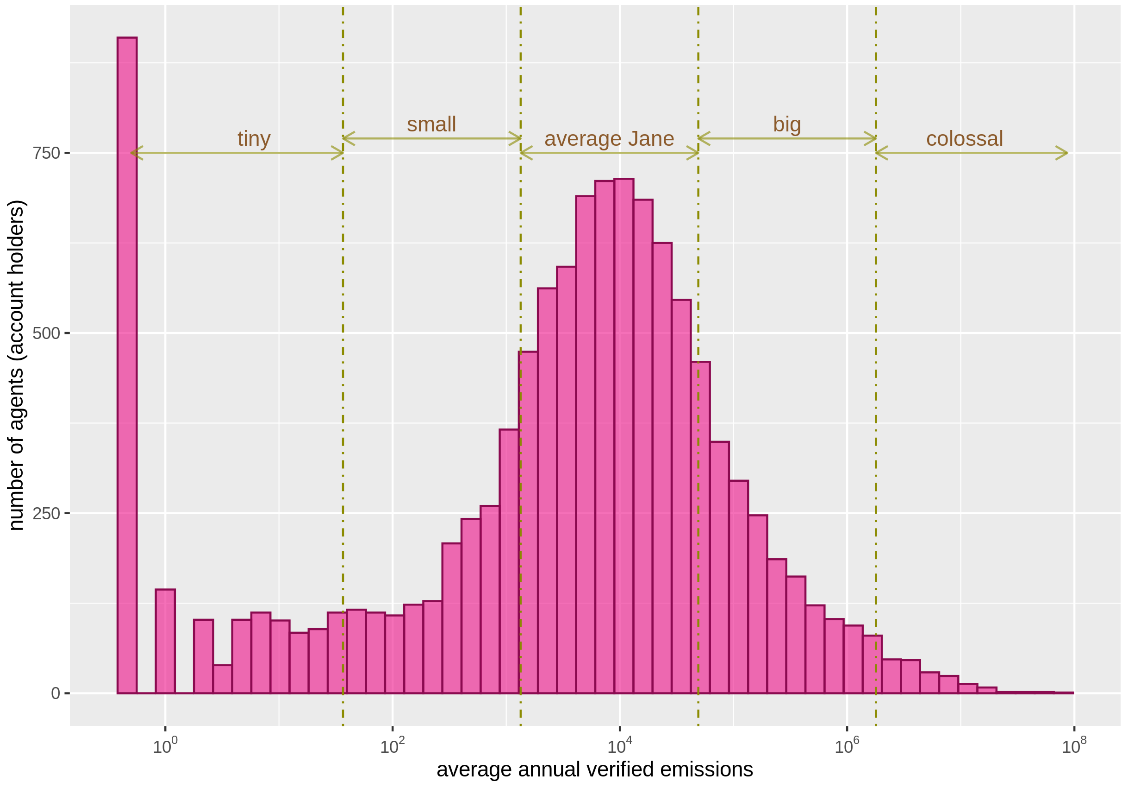


Figure 3.1: Histogram of the average annual emissions, $e_{i;t}$

Chapter 4

Network Analysis

4.1 Preliminaries

In the process of understanding the ETS mechanism we consider the network as graph $G(V, E)$ with V the set of vertices which represent the entities in the system and E the set of edges which represent the transactions between these entities. As explained in the previous chapter there are 3 phases in the range of our dataset, with it reaching the end of 2015 in the midst of the third phase. When we consider the whole graph, there are more than 13000 vertices which causes some problems when processing the network and doesn't produce meaningful results when processed as whole. As we will prove in a later section(4.5), the network exhibits strong periodicity as the processes that define it have a regular occurrence within each year. That inevitably leads to the partition the network to monthly snapshots.

We provided a breakdown of the transaction types based on trade volume and number of transactions in the previous chapter. In most of our analysis and especially in later chapters where we will be interested in constructing models for the network, we will focus on certain types of transaction as they provide the most interesting elements in the system. As seen from the breakdown of transaction types, the lion's share of transactions are of the types, 10-0, 10-2, 3-0, 10-53, 10-36. The main focus of our work is the study of the behavior and structure of the network to those transactions which are allowed a degree of freedom and choice. Considering that 10-2 refer to allowance surrender, 10-53 and 10-36 to allowance allocation they are not choices but obligations and exhibit less interest to the understanding of the entities in the system. As such, we determine, as does Karpf [27] the most essential types to be 10-0, 3-0 and 3-21 the last one being a major type in the first period.

4.2 Degree Assortativity

The concept of assortativity is concerned with the similarity, in terms of some attribute, of connected nodes. Here we are interested in assortative mixing by degree, i.e. how similar the degrees of connected nodes are. A network shows assortative mixing, if high-degree nodes tend to have many connections with other high-degree nodes. There exist many measures for assortativity in literature. We choose to use the Pearson correlation coefficient of degree between pairs of linked nodes, as described by Newman [42].

We define the matrix **e** where the quantity e_{ij} is the fraction of edges in a network that connect a vertex of type i to one of type j. It satisfies the sum rules

$$\sum_{ij} e_{ij} = 1 \; \sum_{j} e_{ij} = a_i \; \sum_{i} e_{ij} = b_i$$

where a_i and b_i are the fraction of each type of end of an edge that is attached to vertices of type i.

We use the Pearson correlation coefficient as defined by Newman [42]:

$$r = \frac{\sum_i e_{ii} - \sum_i a_i b_i}{1 - \sum_i a_i b_i} = \frac{Tr\boldsymbol{e} - \parallel \boldsymbol{e}^2 \parallel}{1 - \parallel \boldsymbol{e}^2 \parallel}$$

where e is the matrix whose elements are e_{ij} and || denoting the 1-norm, that is the sum of the elements of a matrix.

Even though the mixing matrix represents an important metric for the network performance it suffers from the lack of normalization for the number of members of each category.

To ameliorate the situation we provide an assorativity matrix:

$$AM = e - S$$

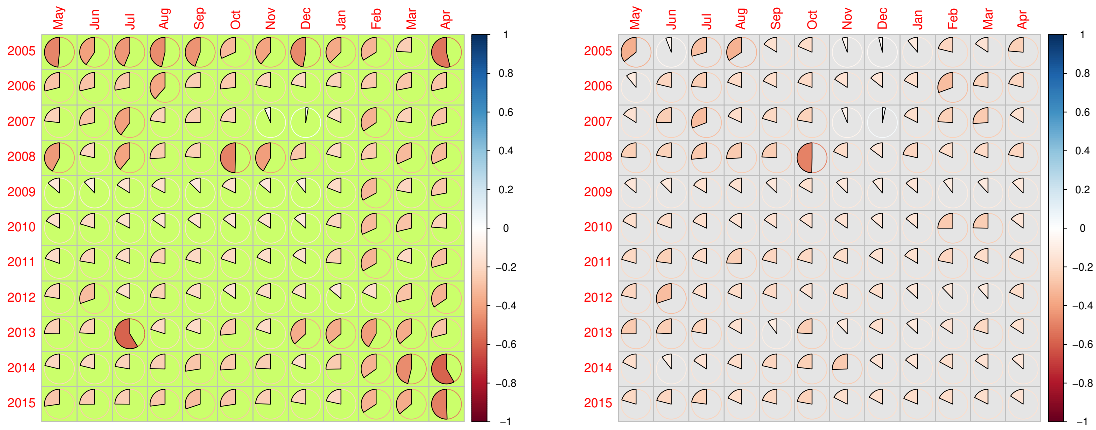
where S is a square matrix consisting of elements $S_{ij} = \frac{\#type_i * \#type_j}{\#edges^2}$

The S matrix is the array of random connections that would occur with the number of members of each category. Therefore the assortativity matrix represents the favorability each category shows to another taking values in (-1,1).

Hence r can take values between -1 and 1 with positive values r indicating a correlation between nodes of similar and 1 corresponding to perfectly assortative(every node connects to a node of the same type of attribute) while -1 to completely disassortative(no same type nodes are connected with an edge).

From what we have observed in social and economic networks as well as citation networks and as analyzed by Piraveenan [46] degree assortativity mainly ranges from -0.2 to 0.4. In the 2017 work Fisher et al.(2017) [20] examined 88 networks of various types including social, transport, biological, mechanical and reported their sizes and their degree assortativity. For the 45 reported socials(with network size larger than 100) the mean degree assortativity is approximately 0.095 with 29 of them having positive assortativity leaving 16 with negative with values ranging from -0.33 to 0.55.

When we examine the values of degree assortativity for the EU ETS network all transactions included, and keeping only transactions types 3-0,3-21,10-0 we get the results as shown in the degree assortativity figure(4.1).



(a) All transactions

(b) Transactions 10-0,3-0,3-21

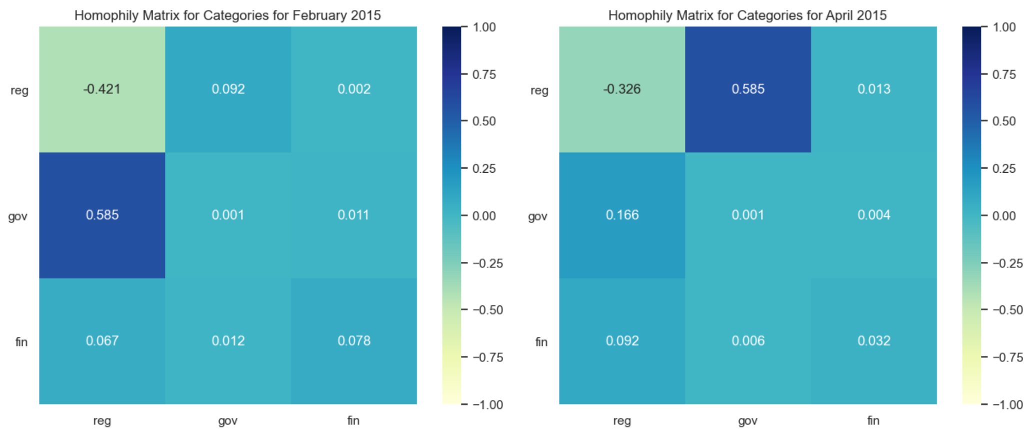
Figure 4.1: Degree assortativity

The Pearson correlation coefficient calculated for each month of the years 2005-2015, showing clearly negative values meaning dissassortativity in terms of degree.

It becomes clear that throughout years and months the Pearson degree correlation has negative values making the network clearly dissassortative. This comes in contrast with results we have found in the literature of social networks [45]. The results we obtain are very much in line with what our experience with the network has been and the snapshots that are provided in the appendix(8). When considering all transactions especially on the months where the major network functions occur, the largest degree nodes are the governmental which either receive surrendered allowances from regulated entities most of whom register single digit transactions, or allocate allowances to those regulated entities.

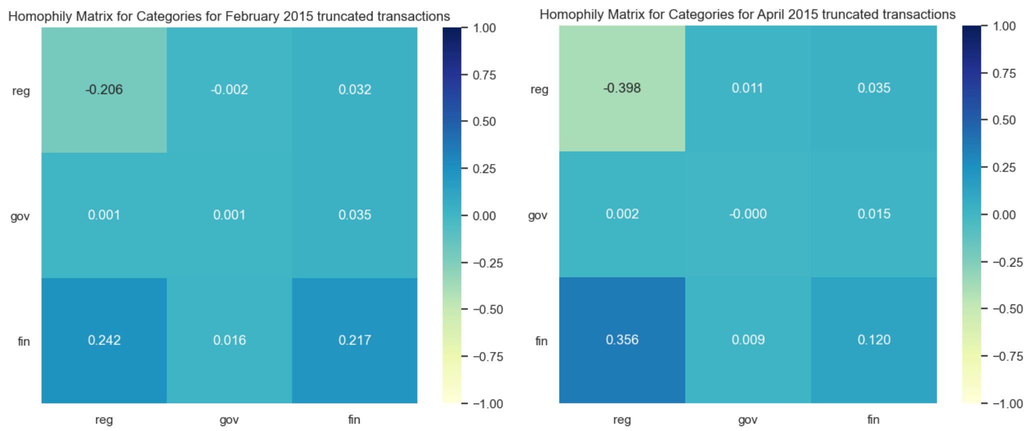
4.3 Category Assortativity

Continuing the methodology and thought process of the previous section we look at the relationship between different types of categories. For the starting months of every year a reasonable expectation is an increased traffic between regulated and governmental entities part of the allocation and surrender of allowances as it becomes obvious in the following figures.



(a) February 2015 (b) April 2015

Considering all transactions we show two key months, February and April.



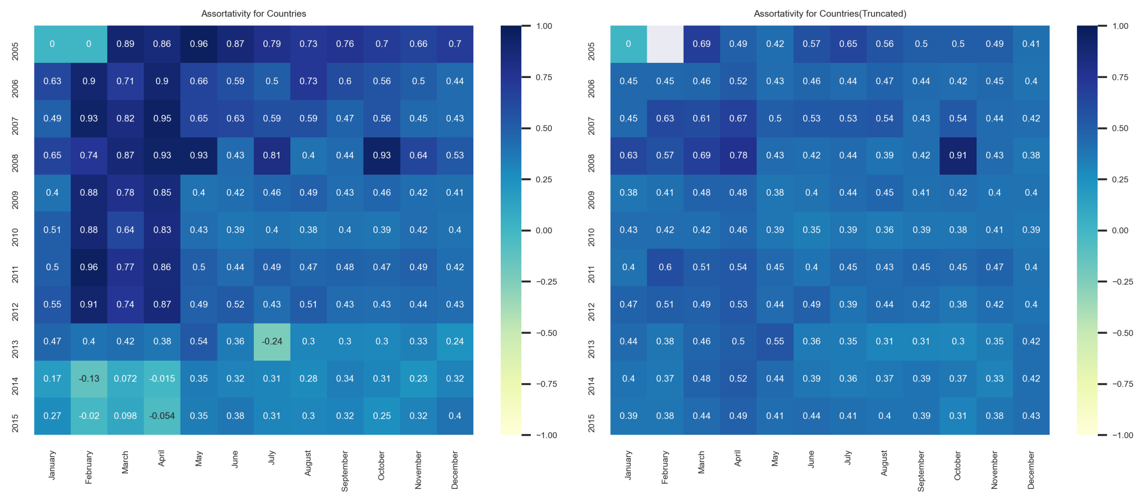
(a) February 2015 (b) April 2015

Considering commercial transactions(3-0,3-21,10-0) we show two key months, February and April.

4.4 Locality

In the effort of trying to understand the reasons EU ETS entities make the choices they make, we study the relationship between their geographic location and the transactions they make. A reasonable assumption to make is that proximity plays some role in the other entities they make transactions and create edges in the network. In the first two phases(2005-2007 and 2008-2012) of the EU ETS the member states were responsible for allocating the allowances. While they didn't seize to exist as entities and keep handling transactions of other kinds, in the third (2013-2020) the European Commission inherited the responsibility of allocating and surrendering the allowances. Since our categorization, as far as the country registry is concerned, includes 'EU' as a country it is expected to produce a significant change in results in the assortativity coefficient for countries.

In figure 4.4, the change from second to third phase becomes visible when looking at the whole of transactions. However when isolating the desired types, we see the positive assortativity coefficient remain pretty much constant throughout the years and phases. What is clear and will be a recurring theme in this thesis, is the peculiar behavior of October 2008, which is distinctly dissimilar to that of other years. This of course is one of the many effects of the financial crisis of that year, which in October saw most of the world's stock exchanges experience the worst declines in their history with drops of around 10% in their indices. It becomes apparent that when we isolate the transactions that form the "commercial" part of the network, it exhibits a more uniform assortativity coefficient. That is to expected as we have removed the most essential component of the mechanism, which is the allocation and surrender of allowances, which in the case of the first years of the system was handled by the member states driving up the assortativity coefficient.



(a) All transactions

(b) Transactions 10-2,3-0,3-21

Figure 4.4: Assortativity Coefficient for Countries

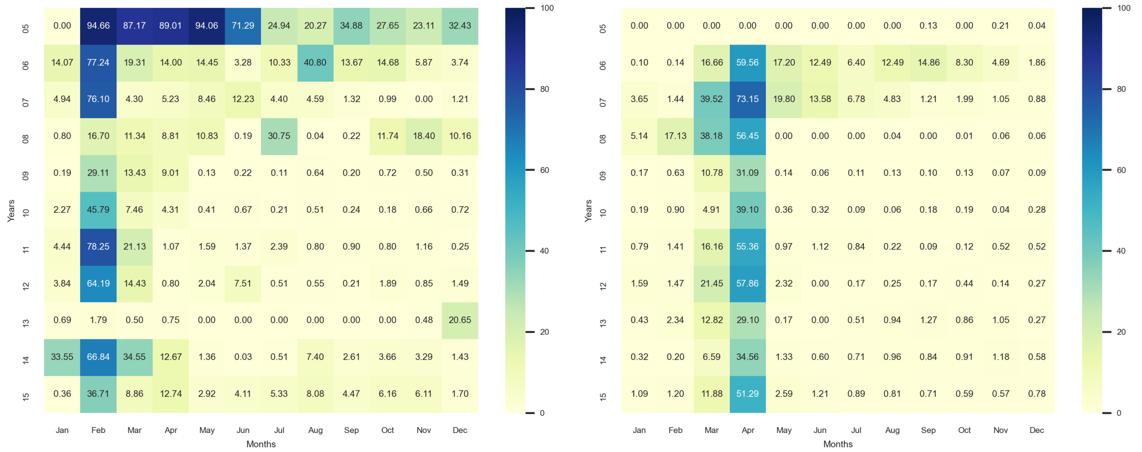
4.5 Periodicity

The essence of the EU ETS mechanism is the surrender of the allowances for each company with the quantity depending on the emissions they show up until the moment they are forced to pay which is the 30th of April every year. The issuing of the allowances to the regulated entities is done at 28 of February every year. The fact that these dates remain constant for the duration of the EU ETS so far, implies that there may be a periodicity in of phenomena happening in these months and the absence of specific EU regulated occasions may imply periodicity in the rest of the year.

In the following figures we show the percentage of transactions of each month that belong to the category mentioned in the caption. By looking at the types of transactions as stated in the figure, those assigned with the tags 10-53 and 10-36 constitute the majority of the transactions in which allowances are allocated to the regulated entities depending on the amount of GHG emissions they had in the past, while those assigned 10-2 are meant for surrender at the end of April.

In the next figure it appears we observe a randomness in the first year of the system and in 2008 which is an anomaly in many regards, but afterwards it becomes clear that the allocation of allowances takes place predominantly in February and the months surrounding it.

In the figure after that it is evident that April is the main month for surrendering allowances which is consistent with the deadline state above while March retains some traffic on these types of transactions most likely due to early surrenders.



(a) % of Transactions in Allocation of Allowances (b) % of Transactions in Surrender of Allowances

Figure 4.5: Pattern of Periodicity

Moreover, in the figures included in the appendix that pertain to the comparison of networks using different methods, it appears that the months of February, March, April and December seem to exhibit unique characteristics while the intermediate months assume a similar form, not having an important role to fulfill in the scheme of the European Transaction System.

4.6 Centrality

The intimate knowledge of the peculiarities of the ETS comes with some insights that facilitate the identification of key characteristics in the network. As we've stated previously the governmental entities are responsible for the dealing and collecting of allowances, the regulated are legally obligated to submit allowances at the end of a period, leaving them the freedom to transact as they please in the meantime and the financial entities fall to neither of the above, and merely seek to make a profit in the system. Considering the nature of the aforementioned types of entities and given the results on the section of the periodicity, for most months the governmental entities do not interact with the entities in the system, while in key months, such as December through April they serve a large and probably central role. In this section we will provide a clear picture of the place each type of entity has in the graph.

To calculate the importance of a node in the graph and its placement we invoke the betweenness centrality formula as state by Brandes [8]:

$$c_b(v) = \sum_{s,t \in V} \frac{\sigma(s,t|v)}{\sigma(s,t)}$$

Denoting by $\sigma(s,t)$ the number of shortest (s,t) paths(sometimes called geodesics) and let $\sigma(s,t|v)$ be the number of shortest (s,t) paths passing through some vertex v other than s,t. If $s = t$, let $\sigma(s,t)$, and if $v \in \{s,t\}$, let $\sigma(s,t|v) = 0$. The measure is therefore usually interpreted as the degree to which a vertex has control over pair-wise connections between other vertices, based on the assumption that the importance of connections is equally divided among all shortest paths for each pair.

In the next figure, we isolate the upper 25% of centralities of 2014, confined to the transactions 10-0, 3-0, 3-21. For the lowest centrality c_{min} of the selected part: $\theta * c_{min} = 1 \rightarrow \theta = \frac{1}{c_{min}}$

For every centrality c of this part: $c' = c{*}\theta$. We graph the quantities $log(c') = log(c{\cdot}\theta)$. We emphasize that we focus on the qualitative properties of the figure not the numeric result.

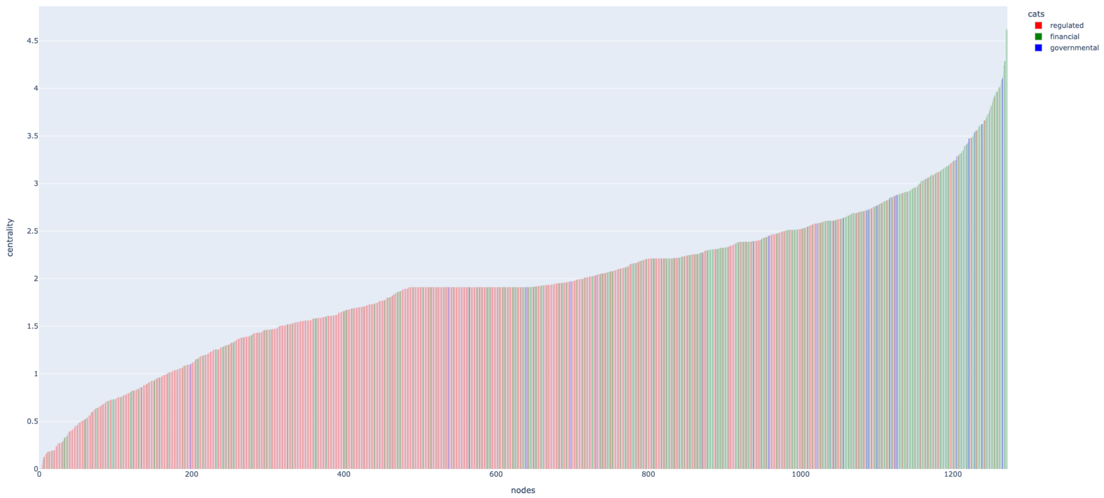


Figure 4.6: Centrality for the 2014 year(truncated) based on category

When not including the surrenders and allocations it becomes evident that the financial nodes and the few governmental ones, which are the auction houses, become the core of the network while the regulated are relegated to the lesser centrality status or even the periphery.

In chapter 2 we introduced a categorization "size" based on the total number of transactions each entity performs. In furthering the knowledge we've obtained for the categories of the network regarding their centrality we are going to attempt the same for the size.

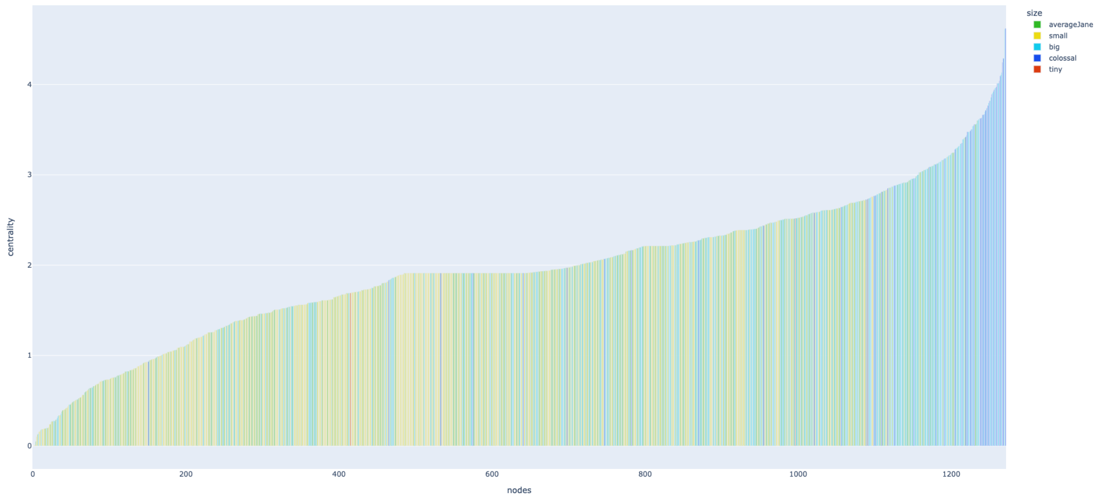


Figure 4.7: Centrality for the 2014 year(truncated) based on size

4.7 Power Law

Across many scientific domains a common claim we see is that many or all real-world networks are scale free. In general a network is characterized as scale free if the fraction of nodes p_k with degree k follows a power-law distribution $k^{-\alpha}$ where $\alpha > 1$. Some versions have stronger requirements e.g. requiring $2 < \alpha < 3$ or that nodes evolve by the preferential attachment mechanism. The universality of scale-free networks, however, remains controversial. Many studies find support for their ubiquity while others challenge it on statistical or theoretical grounds.

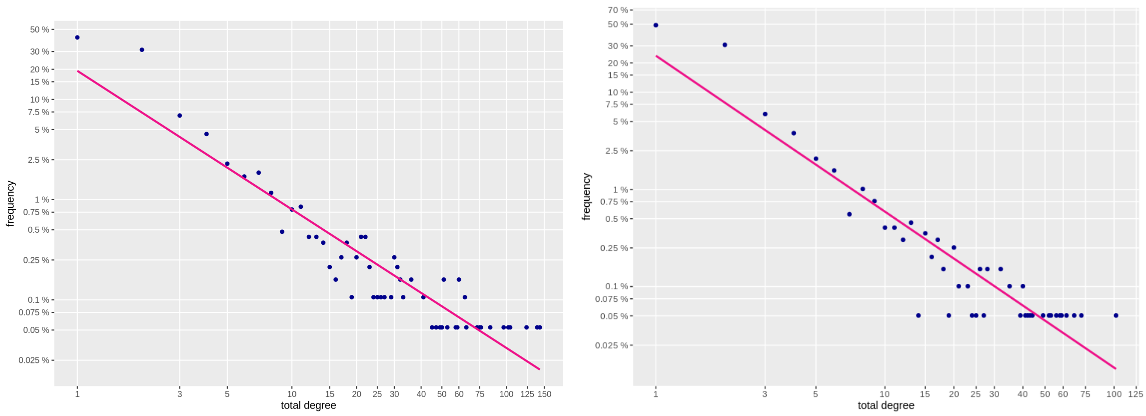
In their 2020 paper Artico et al. [2] devise a statistical testing procedure to test among a sample of 4482 empirical networks the percentage of degree distributions that behave like the tail of a Price model, a two parameter power-law distribution. Their response is to the claim that power-law distribution are rare and the what they find is approximately 64% of the networks as power-law and most of them as scale free. That means the power-law networks not only are not rare but are prevalent.

The methods for testing the existence of power law distribution and scale free networks are ever growing and increasing in precision and concreteness. In their 2019 study Broido et al. [9] devise a test in which they make categorization in 6 terms, based on criteria they meet, not scale free, super-weak, weakest, weak, strong, or strongest.Based on these tests they report that out of 1457 networks, 456 are not scale free, 431 are classified as super-weak, 268 as weakest, 177 as weak, 89 as strong and 36 as strongest. More recently, Artico et al. [2] construct a more robust with the stringent null hypothesis that the network is drawn from an extended de Solla Price network model. In their results they test 4482 real networks and report that for approximately 64% of them they have sufficient power, as power-law most of those are scale free. Considering their stringent hypothesis, using larger power-law class as null the results would be even more one-sided. Even though the tests are many and they range from simple to thorough, we start with a linear regression fitting to gauge the network.

Not all nodes in a network have the same number of edges. The spread in the number of edges a node has, or node degree, is characterized by a distribution function P(k), which gives the probability that a randomly selected node has exactly k edges. The general form of power law is $P(k) \propto k^{-\gamma}$.

$$P(k) = ak^{-\gamma} \Rightarrow \log P(k) = \log a - \gamma \log k$$

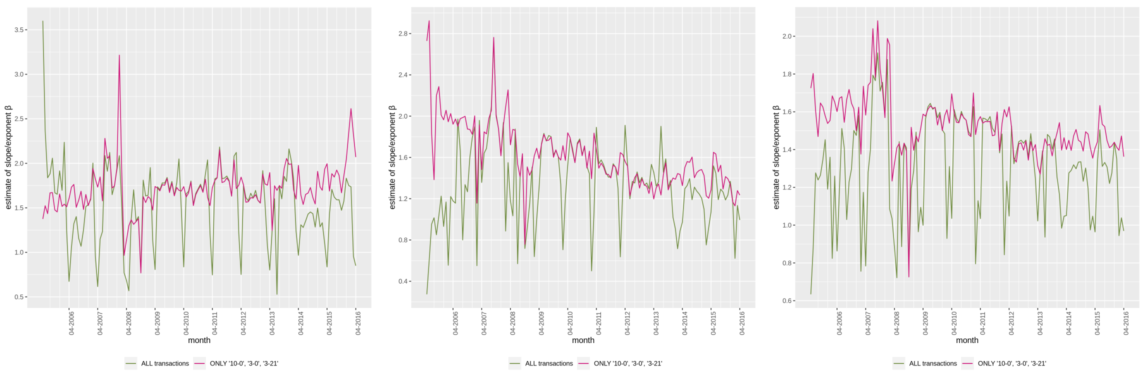
Having P(k) and k data from the network, we run linear regression with parameters $\log a$ and γ serving as intercept and slope respectively. The results for p-value, R^2 and standard error are to be found in the appendix. We test the network per month, while including all transactions and while having truncated them to include only 10-0, 3-21, 3-0 with the results showing in the following figures.



(a) December 2011

(b) April 2012

Figure 4.8: Degree Distributions

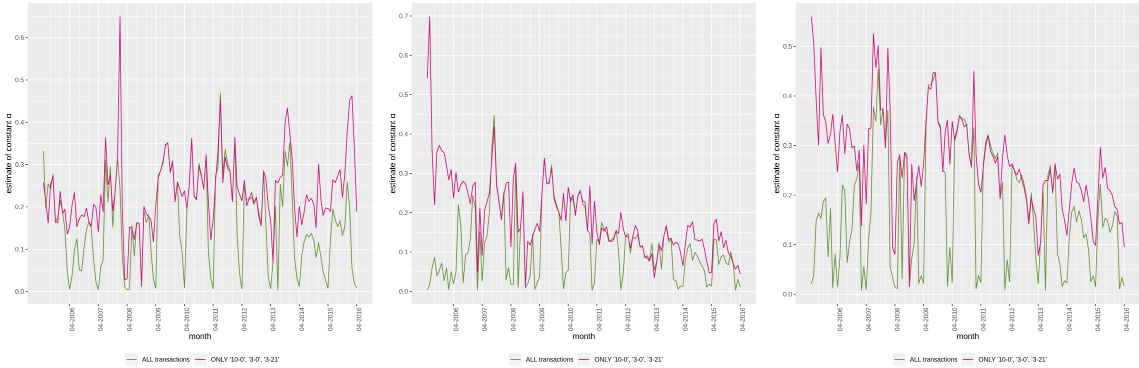


(a) In-degree exponent

(b) Out-degree exponent

(c) Total degree exponent

Figure 4.9: Exponent in linear regression



(a) In-degree constant (b) Out-degree constant (c) Total degree constant

Figure 4.10: Constant in linear regression

It is important to note that as part of this thesis we tested the fit of the power-law in the degree distribution of the monthly networks, but we did not perform comparative analysis trying to determine the absolute best fit, whether it be exponential, lognormal,etc. We consider it beyond the scope of this inquiry as it does not provide any further benefit to the modeling of the networks.

4.8 Densification

All large real-world networks evolve over time by the addition and deletion of nodes and edges. Many models along with shrinking of the effective diameter, exhibit a steady densification over time with the average degree increasing as the number of edges grows superlinearly with the number of vertices. Morevoer, the densification tends to follow a power law pattern. Leskovec et al. [37] find that as the graphs evolve over time, they follow a version of the relation

$$e(t) \propto n(t)^{\alpha}$$

where e(t) and n(t) denote the edges and nodes respectively at moment t, with α the exponent generally taking values between 1 and 2. In that note we examine whether the exponent in the EU ETS network is within those bounds and the implications of the results.

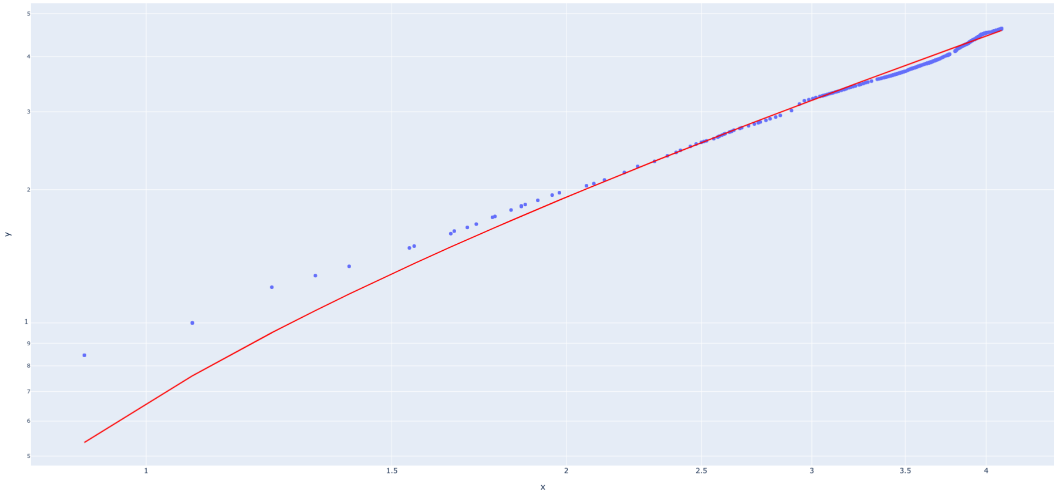


Figure 4.11: Densification for years(2005-2016)

Applying linear regression to the log-edges and log-nodes we get exponent $\alpha = 1.26347925$ which is in line with most networks performance $1 \leq \alpha \leq 2$

4.9 Shrinking Diameters

As networks grow over time, intuition tends to indicate a steady rate of growth for the diameter of the graph. As nodes arrive, edges are created and the expansion of the network ensues. Motivated to test that conventional wisdom Leskovec et al.in their 2007 paper [37] they explored the temporal evolution of diameters expecting to arrive at the point to distinguish between logarithmic and sub-logarithmic growth. Thus it came to some surprise when the diameters in the networks they tested were actually decreasing over time.

To continue with the examination of this hypothesis we have to define some necessary concepts to preempt our observations. Two nodes in a network are connected if there is an undirected path between them. For each natural number d, let g(d) denote the fraction of connected node pairs whose shortest connecting path has length at most d. Graph G has diameter d if the maximum length of undirected shortest path over all connected pairs of nodes is d. Directionality will not be taken into account when studying the diameters or effective diameters as will be described.

Effective Diameter For each natural number d, let g(d) denote the fraction of connected node pairs whose undirected shortest connecting path in a graph G has length at most d. And let D be an integer for which $g(D-1) < 0.9$ and $g(D) \geq 0.9$. Then the graph G has the integer effective diameter D. In other words, the integer effective diameter is the smallest number of hops D at which at least 90% of all connected pairs of nodes can be reached.

They also extend that function to include real values of x for g(x). By linearly interpolating the function value between $g(d)$ and $g(d+1)$ ($d \leq x < d+1$) we get:

$$g(x) = g(d) + (g(d+1) - g(d))(x - d)$$

An alternative definition: Let D be a value where $g(D) = 0.9$. Then graph G has the effective diameter D. This definition varies slightly from an alternate definition of the effective diameter used in earlier work: the minimum integer value d such that at least 90% of the connected node pairs are at distance at most d. This variation smooths this definition by allowing it to take noninteger values.

The use of effective diameter comes from its robustness compared to the diameter(defined as the maximum distance over all connected node pairs) since the diameter is prone to effects of degenerate structures in the graph e.g very long chains. In their 2007 paper Leskovec et al. [37] state that their experiments show that the diameter and effective diameter exhibit a qualitatively similar behavior, although they don't state that this be the case universally.

As effective diameter is a superior metric in the following figure we exhibit the effective diameters of the network taken at intervals of 5 days, showing a clear descending pattern which is highlighted by plotting the ordinary least squares regression line.

Showing in the following figure is the effective diameter evolution over time, for both the whole network and the truncated(3-0, 3-21, 10-0). It is evident that both follow a downward slope of -0.0017 and -0.002 for whole and truncated respectively.

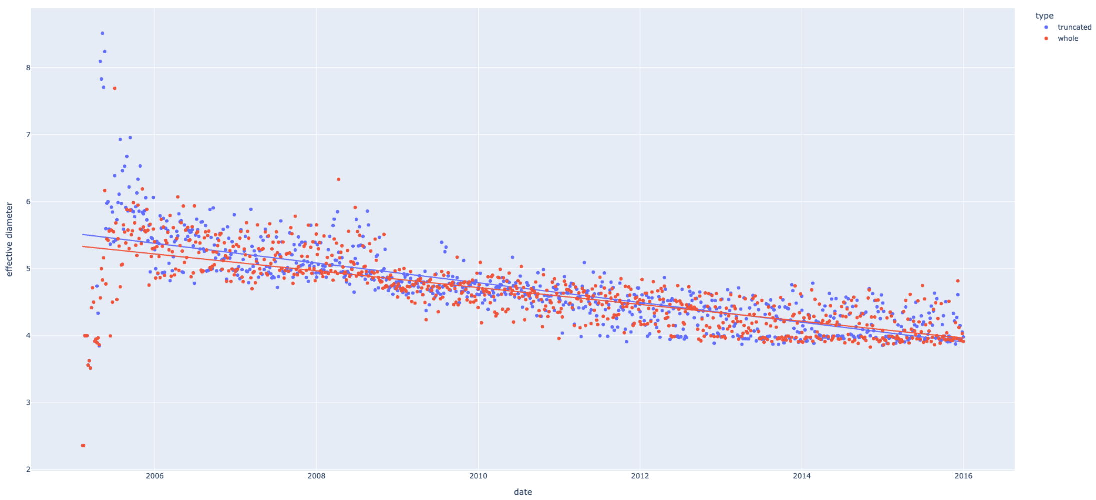


Figure 4.12: Effective Diameter of time

Our analysis in the whole of the thesis will be largely focused on monthly static networks. As such we are interested to see if we derive the same behavior in a shorter period of time with significantly fewer nodes. We have tested every month for the years 2005-2015(January of 2005 had no transactions) for the truncated transactions we have described for the evolution of the effective diameter over the days. When running linear regression on the values for those months we come to the following results. In 61 of those months we see a rising effective diameter and in the other 70 we see it shrinking, which is remarkable considering the much smaller scale we operate on. This is probably because of the fact they have the tendency to attach to already popular nodes, either being governmental or financial.

Chapter 5

Comparison Measures

A difficult question when studying networks is identifying similar ones through comparison. When provided with real-world networks we have to be able to determine whether a stochastic network model can accurately. We reviewed several measures to find a suitable one for the purposes of this thesis. We first looked at DeltaCon by Koutra et al [30], a scalable metric based on node affinity, which works very well in Known Node Correspondence networks(networks of the same node set) but needs modification to be of use to us in the context of this thesis.

In the area of Unknown Node Correspondence(UNC) there are several alignment-based methods like the GRAAL family [32] with its variations(L-[38], MI-[33], C-[39], H-[40]) which create a mapping between the nodes of two graphs (alignment procedure) trying to maximize an objective function that captures the quality of matching. A different approach is considering graphlets. They are small, connected, non-isomorphic subgraphs of large networks. They encode important information about the structure of the network and provide a valuable tool for comparison. The work done by Yaveroglou [54] covers a broad range of metrics and ways to use graphlets to depict and compare networks such as Graphlets Correlation Distance (GCD), Graphlet Degree Distribution Agreement (GDDA) and Relative Graphlets Frequency Distance (RGFD).

While many of these options are reliable metrics, we decided to use Portrait Divergence as described by Bagrow in his 2019 work [3]. Our choice, which will become clearer with the description of the process, was based on the ability to distinguish between different types of structure, being a UNC metric, the inclusion of direction and/or weights if decided and the fact that there was reliable code by the author. In this chapter we will define the notion of the portrait and how it will be used to compare networks.

5.1 Portrait Divergence

In order to define the Portrait Divergence metric we need to introduce the concept of the B-matrix [4]. The distance between two nodes u,v is the smallest number of edges between the nodes which can be found with Breadth First Search(BFS). Considering a node v_i, an l-shell is the set $V_l \subseteq V$ of nodes at distance l from v_i. So it is stated:

$B_{l,k}$ = number of nodes that have exactly k members in their respective l-shells

From the definition above it can be surmised that:

$B_{l,k} = NP_l(k)$

where N is the number of nodes and $P_l(k)$ the percentage of nodes at degree of order l

Apart from providing an impression of the form of the network the portrait will be used to compare networks G and G'. Taking inspiration from the syllogism behind the Kolmogorov-Smirnov test, they define the following statistic for corresponding rows B_l and $B_{l'}$:

$$K_l = max_k |C_{l,k} - C'_{l,k}|$$

where C is the matrix of cumulative distribution of B:

$$C_{l,k} = (\sum_{k' \leq k} B_{l,k'}) / \sum_{k'} B_{l,k'}$$

The statistic defines a two-sample hypothesis test for whether or not the corresponding rows of the portraits are drawn from the same underlying, unspecified distribution. In the case that the two graphs have different diameters, the portrait of the smaller diameter graph is expanded to the same size as the larger by defining empty shells $l < d$ as $B_{l,k} = N\delta_{0,k}$.

Finally we define the scalar distance $\Delta(G, G')$:

$$\Delta(G, G') \equiv \Delta(B, B') = \frac{\sum_l a_l K_l}{\sum_l a_l}$$

where

$$a_l = \sum_{k>0} B_{l,k} + \sum_{k>0} B'_{l,k}$$

is a weight chosen to increase the impact of the lower, more heavily occupied shells

This doesn't comprise the Network Portrait Divergence metric as Bagrow et al. define it in their 2019 work [3]. In that work the describe an information theoretic measure, with a number of desirable properties compared to the ad hoc one above.

The rows of B may be interpreted as probability distributions:

$$P(k|l) = \frac{1}{N} B_{l,k}$$

is the (empirical) probability that a randomly chosen node will have k nodes at distance l. An immediate comparison per row arises:

$$KL(P(k|l)||Q(k|l)) = \sum_k P(k|l)) = \sum_k P(k|l)log\frac{P(k|l)}{Q(k|l)}$$

where $KL(p||q)$ is the Kullback-Liebler (KL) divergence between two distributions p and q, with Q defined as $P(k|l)$ as defined above but for the second portrait. They admit that it suffers from some drawbacks such as, undefined $KL(P(k|l)||Q(k|l))$ in some cases and lack of symmetry therefore not a distance.

In their effort to fix them they provide the KL-divergence as stated below.

$$KL(P(k|l)||Q(k|l)) = \sum_{l=0}^{max(d,d')} \sum_{k=0}^{N} P(k,l)log\frac{P(k|l)}{Q(k|l)}$$

Thus the **Network Portrait Divergence** is

$$D_{JS}(G,G') = \frac{1}{2}KL(P||M) + \frac{1}{2}KL(Q||M)$$

with $M = \frac{1}{2}(P+Q)$being the mixture distribution of P and Q. Here P and Q are defined from $P(k,l) = \frac{kB_{l,k}}{\sum_c n_c^2}$ as stated in their paper.

The Network Portrait Divergence $0 \leq D_{JS} \leq 1$ provides a single value to quantify the dissimilarity of the two networks by means of their distance distributions, with smaller D_{JS} for more similar networks and larger D_{JS} for less similar networks

5.1.1 Weighted

The original work defining network portraits (Bagrow et al. 2008) did not consider weighted networks, where a scalar quantity w_{ij} is associated with each $(i,j) \in E$. An important consideration is that path lengths for weighted networks are generally computed by summing edge weights along a path, leading to path lengths $l \in R$ (typically) instead of path lengths $l \in Z$. To address this, we define in more detail the construction of the weighted portrait divergence.

To generalize the portrait of weighted networks requires using an algorithm for finding shortest paths accounting for edge weights(Djikstra will be used) and defining an appropriate aggregation strategy to group shortest paths by length to form the rows of B.

To aggregate shortest paths by length is to introduce a binning strategy for the continuous path lengths. Let $d_0 = 0 < d_1 < \cdots < d_{b+1} = L_{max}$ define a set of b intervals, where L_{max} is the length of the longest shortest path. The weighted portrait B can be defined such that $B_{i,k} \equiv$ the number of nodes with k nodes at distances $d_i \leq l < d_{i+1}$. So, the i-th row of the weighted portrait accounts for all shortest paths with lengths falling inside the i-th bin $[d_i, d_{i+1})$.

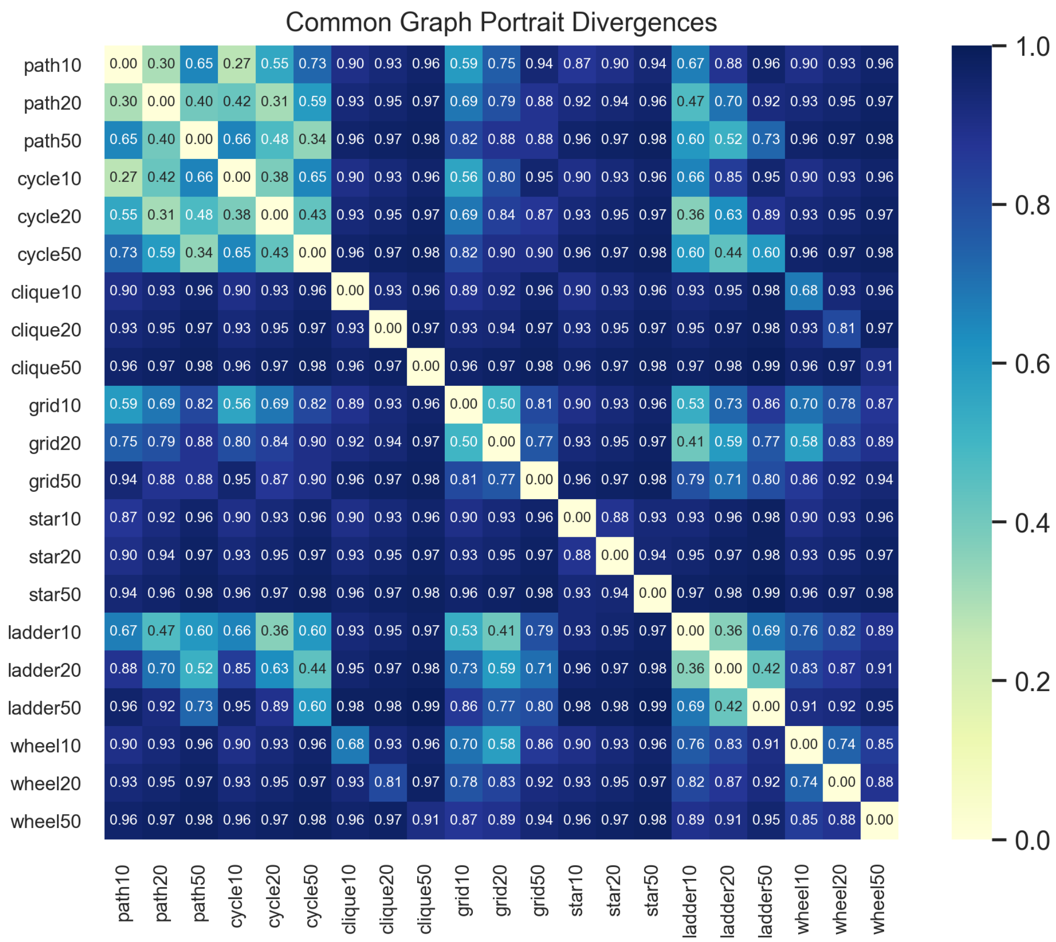
5.1.2 Notes on Portrait Divergence

At this point, it is imperative to note that, as noted by the authors in their 2019 work [3] on the Portrait Divergence metric, there has yet to be conducted serious benchmarking for their metric apart from some results they present for a couple simple synthetic networks and some real world examples. As such, to provide an accurate assessment of the performance of the network building algorithms that will be presented in following chapters we have to compare it against the reference data we have collected on the comparison of each network snapshot to the others.

The usage of this metric, apart from its interesting and useful network portrayal, is motivated heavily by its embankment on the UNC(Uknown Node Correspondence) category, meaning that it states that the networks for comparison need not have the same node set V. As we have chosen to look at the network in monthly snapshots, the node set changes drastically from month to month leaving us no option to use any Known Node Correspondence methods.

An important feature which Tantardini et al [49] remarked in their conclusion of their 2019 comparison of various methods of network comparison, is the structural comparison it manages to achieve, namely to provide information about how much, and in what sense, the structures of graphs differ. They also highlight that contrary to other measures they tested, it performs better in the undirected case rather than the directed.

At this point we have to provide a reference for our later results since there are not that many real network benchmarks. Considering that 2014 and 2015 have been some of the most stable years and moving forward with the knowledge that this phase will through 2020, we will exhibit a comparison of all the months of 2014.We first show some results for basic graphs for 10,20 and 50 nodes and we can see clearly that when two graphs are not of the same type their Portrait Divergence is very high.


Common Graph Portrait Divergences
path10
path20
path50
cycle10
cycle20
cycle50
clique10
clique20
clique50
grid10
grid20
grid50
star10
star20
star50
ladder10
ladder20
ladder50
wheel10
wheel20
wheel50
1.0
0.8
0.6
0.4
0.2
0.0

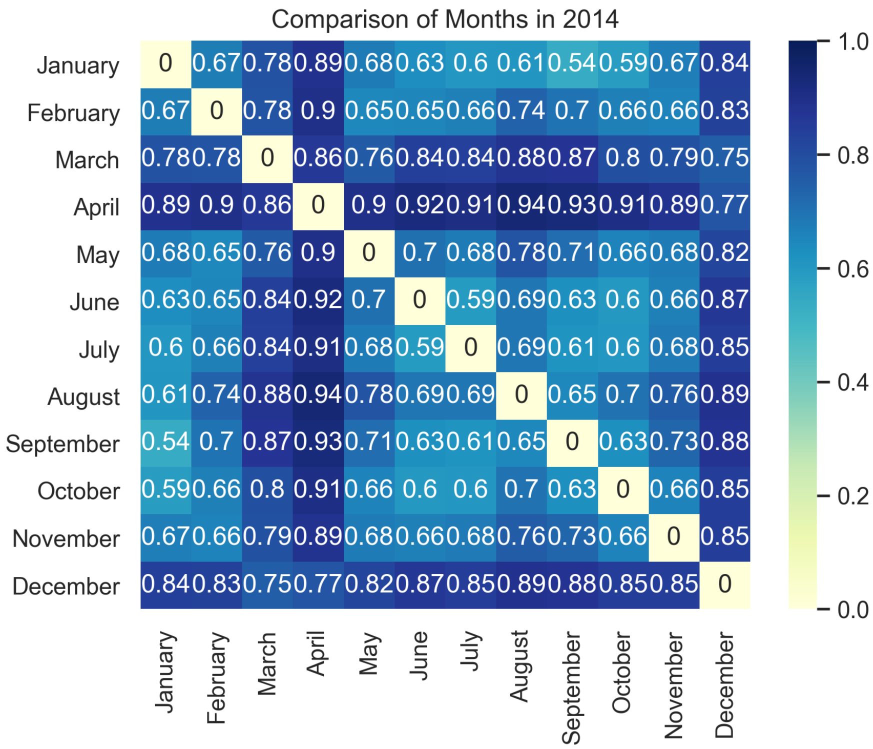


Figure 5.1: Reference(Undirected Unweighted)

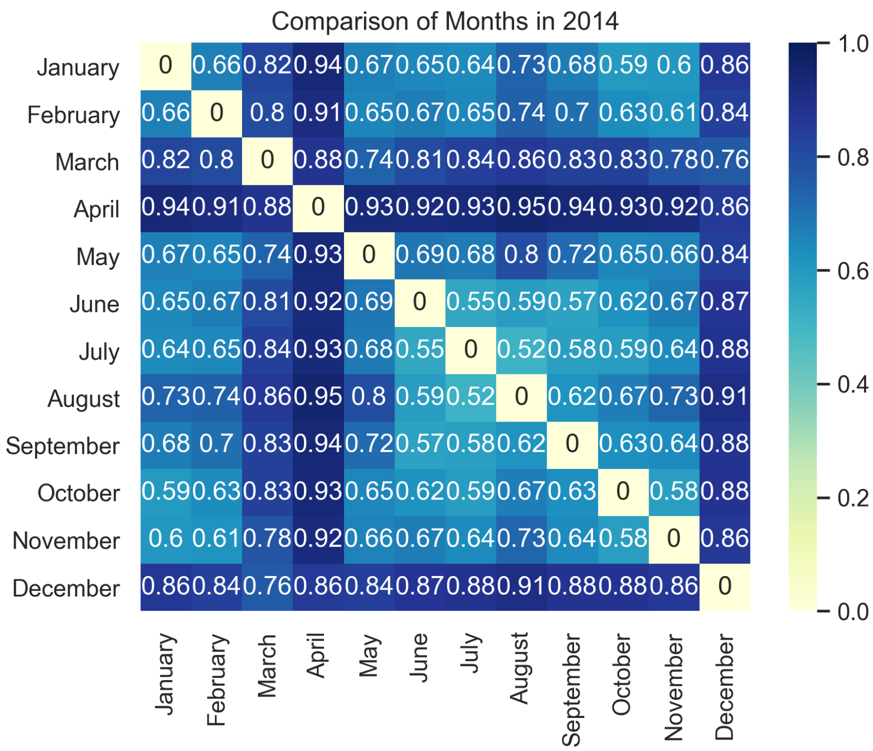


Figure 5.2: Reference(Directed Unweighted)

Chapter 6

Generative Models

As demonstrated in the chapter 3, when examining the effective diameters of the network we find that it decreases over time when accounting for the static network of all the years and shows a decreasing pattern in many(almost half) of the monthly static networks. Confirming this counterintuitive hypothesis leads us further into the work of Leskovec et al. [37] who tried to create models to replicate the behavior in the networks they studied. In their search for models that recreate real-life networks that follow power law in terms of degree distribution or exhibit power law like characteristics they created some models that we applied to approach the EU ETS model.

6.1 Community Guided Agreement

This approach is motivated by the realization that power laws appear in combination with self-similar structures. Intuitively, a self-similar object consists of miniature replicas of itself. The first model they created and the first we will try is the Community Guided Agreement. We represent the recursive structure of communities-within-communities as a tree Γ of height H. Leskovec et al. show that even a simple perfectly balanced tree of constant fanout b is enough to lead to a densification power law. The nodes V of the graph will be the leaves of the tree($n = |V|$ and $n = b^H$). Let $h(v, w)$ define the standard tree distance of two leaf nodes v and w, that is, $h(v, w)$ is the height of their least common ancestor(the height of the smallest subtree containing both v and w).

We construct a random graph on a set of nodes V by specifying the probability that v and w form an edge as a function f of $h(w, v)$. That will be referred to as the difficulty function. It becomes evident that this should decrease with h, but there are many forms such a decrease could take.

The form of f they decide works best comes from the self-similarity arguments they made. It should be scale-free, so $\frac{f(h)}{f(h-1)}$ should be level independent and thus constant. That implies that the definition should be $f(h) = f(0)c^{-h}$. If for simplicity $f(0)$ is set to 1, then:

$$f(h) = c^{-h}$$

where $c \geq 1$ and c is referred to as the difficulty constant. Intuitively, cross-communities links become harder to form as c increases.

6.2 Forest Fire Model

In order to capture other characteristics which the CGA couldn't, they created the Forest Fire Model. In particular, they wanted to capture both the shrinking effective diameters that they observed as well as the fact that real networks tend to have heavy-tailed out-degree distributions (though generally not as skewed as their in-degree distributions).

Nodes arrive one at a time and form outlinks to some subset of the earlier nodes. To form outlinks, a new node v attaches to a node w in the existing graph, and then begins burning links outward from w, linking with a certain probability to any new node it discovers. Formalizing the process:

We begin with two parameters, a forward burning probability p and a backward probability ratio r whose roles will be described in the following. When we consider a node v joining at time $t \geq 1$, and let G_t be the graph constructed thus far. The node v forms outlinks to nodes in G_t according to the following process.

- v chooses an ambassador node w uniformly at random and forms a link to w
- generate two numbers, x and y, that are geometrically distributed with means $\frac{p}{1-p}$ and $\frac{rp}{1-rp}$ respectively. Node v selects x out-links and y in-links of w incident to nodes that were not yet visited. Let $w_1, w_2, \cdot, wx + y$ denote the other ends of these selected links. If there are not enough in-links or out-links available , v selects as many as it can.
- out-links are formed between v and each of $w_1, w_2, \cdot, w_{x+y}$. As the process continues, nodes cannot be visited a second time, preventing the construction of cycling.

Many properties arise from this model and as they have been established through simulation from Leskovec et al. we present some intuition the mechanism of the model provides.

- **Heavy tailed in-degree** The model follows a rich get richer pattern as highly linked nodes can be easily reached by a newcomer regardless of ambassador.
- **Heavy tailed out-degree** Due to the recursive nature of link formation there is a reasonable chance for a new node to burn many edges and thus produce a large out-degree.
- **Densification power law** A newcomer will have a lot of links near the community of his/her ambassador, a few links beyond this, and significantly fewer farther away. Intuitively, this is analogous to the Community Guided Attachment, although without an explicit set of communities.

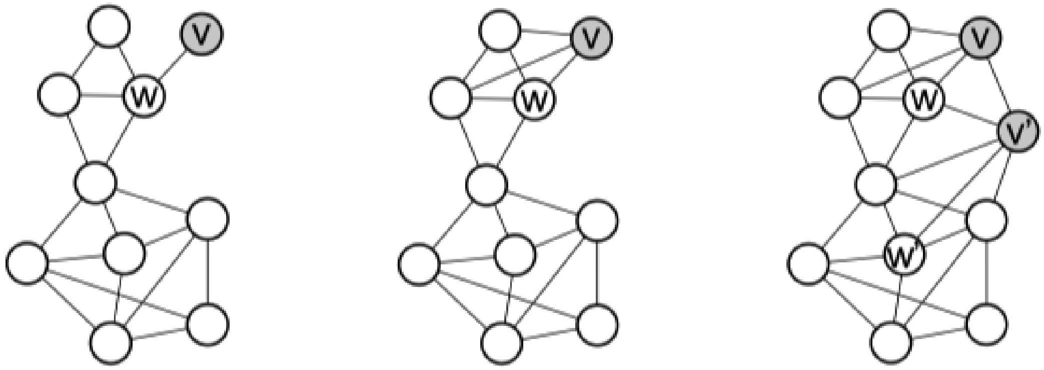


Figure 6.1: The Forest Fire burning process.
Left: a new node v joins the network and selects a seed node w. Middle: v then attaches itself by recursively linking to w's neighbors, w's neighbor-neighbors, and so on, according to the "forest fire" burning mechanism described in the text. Right: a new node v' joins the network, selects seed w' , and recursively adds links using the same "forest fire" burning mechanism. Notice that if v' causes a large "fire" it links to a large number of existing nodes.

- **Shrinking Diameter** Through the mechanism described in the model it isn't evident how a shrinking diameter phenomenon can be accomplished. Graph densification is helpful in reducing the diameter, but it is important to note that densification is certainly not enough on its own to imply shrinking diameter.

6.3 Preferential Attachment

Many social networks are characterized by a highly uneven distribution of links. The observed skewed distributions have in many cases been attributed to preferential attachment, a tendency among nodes in a growing network to form new links preferentially to nodes with high numbers of links. As it has been stated many times previously in the literature, preferential attachment mechanisms generate distributions which are approximately power law over transient periods[31] [19]. As shown in the network analysis chapter, our network follows a power-law degree distribution making it more than reasonable to explore if such models replicate the network's structure.

6.3.1 Price's Model

In 1965 physicist-turned-historian-of-science Derek de Solla Price described what probably is the first example of what we now know as a scale-free network[47]. While studying networks on citations in scientific papers, he observed that in-degrees and out-degrees follow power-law distributions. Like many after him, his work is based on ideas of Herbert

Simon who showed that power laws arise when there exists a "rich get richer" pattern, that is the amount you get increases with the amount you have. Price named this cumulative advantage and, as will later be analyzed, it has become known as *preferential attachment* as coined by Barabasi and Albert. The model:

We consider a directed graph of n vertices and let p_k be the fraction of vertices with in-degree k, such that $\sum_k p_k = 1$. New vertices arrive continuously in the network and every vertex that arrives is assigned a fixed number of out-degree at the creation of the vertex. The out-degree may vary between nodes but the mean degree denoted m is constant over time. As m is a mean it can take non-integer values. The mean in-degree is also m as $\sum_k kp_k = m$.

The probability that a newly appearing vertex,e.g. a new paper cites an existing one, connects to previous vertices is simply proportional to the in-degree k of the old vertex. As each vertex starts out with zero in-degree and hence would forever have zero probability of gaining new edges. To circumvent the problem the probability should be proportional to $k + k_0$. A convention, which can be justified by saying that one can consider the initial publication of a paper to be its first citation, is to set $k0 = 1$. So the probability that a new edge attaches to any of the existing vertices with degree k is:

$$\frac{(k+1)p_k}{\sum_k (k+1)p_k} = \frac{(k+1)p_k}{m+1}$$

As proven in Newman's 2003 paper [43]:

$$p[k] \approx k^{-(2+1/m)}$$

In other words, in the limit of large n, the degree distribution has a power-law tail with exponent $\alpha = 2 + \frac{1}{m}$. This will typically give exponents in the interval between 2 and 3, which is in line with the values seen in the network he examines. The equation above does not depend on k_0 and therefore the offset parameter $k_0 = 1$ can be justified a posteriori.

6.3.2 Barabási-Albert

Price's work remained largely unknown to the scientific community and cumulative advantage did not achieve currency until its discovery by Barabási and Albert [5] who gave it a new name, preferential attachment. The difference between the two models is that in the model of Barabási-Albert edges are undirected so there is no distinction between in and out-degree. This comes with pros and cons, the positive being that by ignoring the directed nature of the network, the model of Barabási and Albert gets around Price's problem of how a paper gets its first citation or a Web site gets its first link. On the other hand we are losing a crucial feature of networks, direction. Looking at it pragmatically, the model of Barabási and Albert is as a model that sacrifices some of the realism of Price's model in favor of simplicity.

The probability that a new edge attaches to a vertex of degree k—the equivalent of the price equation is

$$\frac{kp_k}{\sum_k kp_k} = \frac{kp_k}{2m}$$

As noted by Barabasi himself in his 2013 book [6] there is a number of analytical tools to calculate the degree distribution of Barabasi-Albert model. Using continuum theory he predicts:

$$p(k) \approx 2m^{1/\beta}k^{-\gamma}$$

That β is called dynamical exponent and $\beta = \frac{1}{2}$ and from the continuum theory comes $\gamma = \frac{1}{\beta} + 1 = 3$. So the degree distribution equation becomes:

$$p(k) \approx 2m^2k^{-3}$$

6.4 Comparison Of Performance

Using the portrait divergence metric it is necessary to create subcategories for comparison as directed and undirected network have different a comparative measure in portrait divergence.

6.4.1 Undirected models

In this section structure becomes the foremost characteristic in the recognition of similarity between the model and the real network. The Barabasi-Albert model has been the start of every sentence involving networks for the past 20 years, owing its success to its simplicity and intuitiveness in the approach of the scale-free property. As stated in the previous section the distribution equation is:

$$p(k) \approx 2m^2k^{-3}$$

Through our degree distribution it becomes apparent that we are far from the exponent in the equation above, which is also immutable. The only other modifiable quantity is the constant, which is dependent on m. However considering the results in previous chapter, fitting that parameter to our results would result in an m quantity which produces a far sparser network than ours.

We can still test whether the model can produce a servicable result by fixing the m constant to produce a very similar density. In order to achieve that we have to make some modification to the original BA model. It is impossible to achieve such a result with an integer m. As was the method with the Price model, we allow non integer values and in each turn we will draw m from a normal distribution with m as its mean. Achieving the closest density to the original network we perform our test with m set to $\frac{E(G)}{V(G)} = \overline{k}$ where E(G) the number of edges and V(G) the number of vertices and $\overline{k}$ the mean degree.

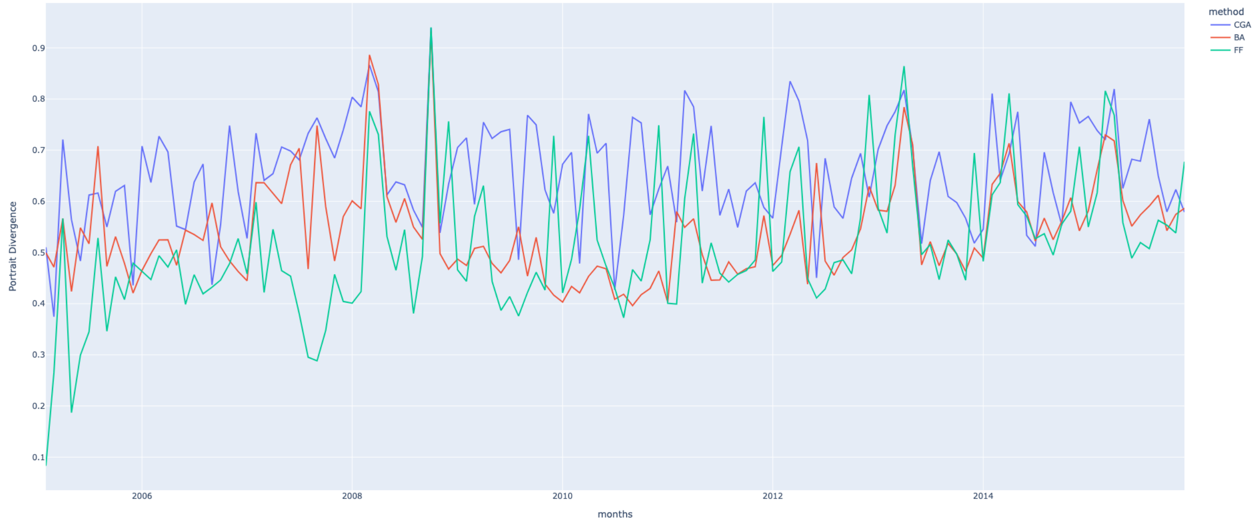


Figure 6.2: Comparison of Undirected Methods

Performance-wise it seems the BA model can capture most accurately the structure of the monthly networks with the Forest Fire also having satisfactory performance despite volatility. The Portrait Divergence results are $\overline{PD}_{BA} = 0.541$, $\overline{PD}_{FF} = 0.515$ with $\overline{p} = 0.335, \overline{r} = 0.271$ and $\overline{PD}_{CGA} = 0.656 (c = 1.4)$.

6.4.2 Directed models

Acquiring an additional dimension to the modeling of the network, directionality necessitates different models and the expectation is different results, in that it will be more difficult to simulate the pattern of direction of actual edges.

As desrcibed above for the Barabasi Albert model, the m input of the Price model will be $m > 2$ and as we have shown we don't have that type of exponent, so we will operate accordingly again here by setting it to $\frac{E(G)}{V(G)} = \overline{k}$.

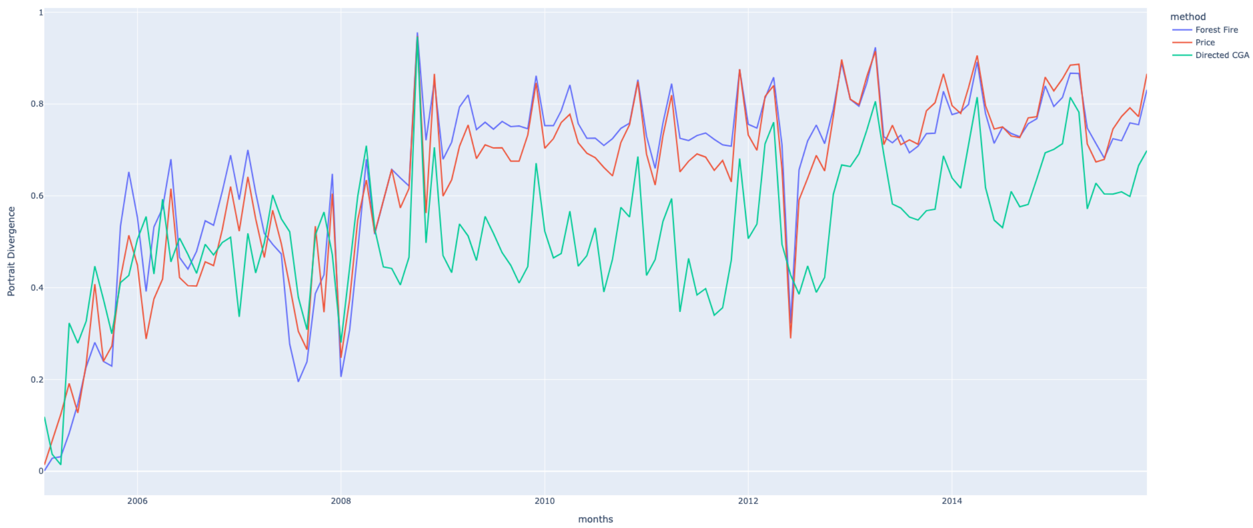


Figure 6.3: Comparison of Directed Methods

In terms of performance, when randomly applying directions to the CGA model it comes closer to the monthly networks than its competitors with $\overline{PD}_{DCGA} = 0.521$, $\overline{PD}_{Price} = 0.634$ and $\overline{PD}_{FF} = 0.653$ with $\overline{p} = 0.245, \overline{r} = 0.403$.

As it becomes apparent in the results for the directed models, the performance when considering direction becomes significantly worse as the stochastic process is somewhat delicate to the addition of directionality dimension. As with the undirected results, it seems that the periodicity we described in Chapter 4 remains ever present as the peaks remain in the most "complex" months, December and April. However unlike the undirected the low complexity and smaller scale of the network in the first period and the start of the second seem to be ripe for modeling with both of the models used.

It is clear that the Forest Fire model performs noticeably worse in the directed case, which can be attributed to the fact the direction of edges in a real network tends to flow mostly outwards as smaller nodes receive allowances from bigger financial or governmental nodes. In reality there is also large reciprocity, which the Forest Fire can't replicate. A particularly interesting observation is the shape and closeness of the two curves, which in some months arrive at almost the same value of Portrait Divergence, showing that the performance of the Forest Fire model empirically follow the power law standard that 30 years prior Price set in his.When testing if similar graphs are recreated by both process we arive at Portrait Divergence values over 0.85 which means both models capture the structure of the network in different ways.

Chapter 7

Conclusion

7.1 Concluding Remarks

To summarize, in this work we provided a thorough analysis of some key network statistics that provide a clear understanding of the structure of the EU ETS network that has not been achieved in the literature. In the second part of the thesis we tested some models in order to approach the static monthly network both directed and undirected, using network comparison measures which we haven't seen done in the literature.

Specifically, we showed that, while in most real world networks there exists a uniformity as to degree assortativity, in EU ETS network nodes of similar degrees tend to not share an edge, confirming the impression we got when studying the structure of the network visually. When examining the behavior of actors considering their categorization, clear patterns emerge as to which transactions they tend to make and as to be expected they show a much increased tendency to transact with other actors of their own country.

The network exhibits strong elements of annual periodicity and predictability as far as the structure and traffic are concerned, despite the change in phases. As it has become commonplace in network analysis, we examined if the network follows a power-law degree distribution and with almost no exception in months it obeys scale-invariance, although with relatively small exponent.

An important realization when studying networks is their evolution over time, which when we examined the network's progress it became evident that it increases in density, which is in line with the literature. We also looked at how its effective diameter evolved over time, and counterintuitively it decreased.

Following the network analysis, we introduce Portrait Divergence, a known comparative measure for networks, preempting the creation of artificial networks with stochastic models. In the undirected case we examined one of the most common ones, the Barabasi-Albert model which we modified to provide a better fit for our case. We compared it to the Community Guided Agreement and the undirected version of the Forest Fire model, though all satisfactory approaches, the BA model recreates the network structure most accurately in the more stable phases. In the directed case, the Price model is put to the test when compared to Forest Fire model and the directed Community Guided Agreement with the latter outshining the others, approaching the structure of the network more closely.

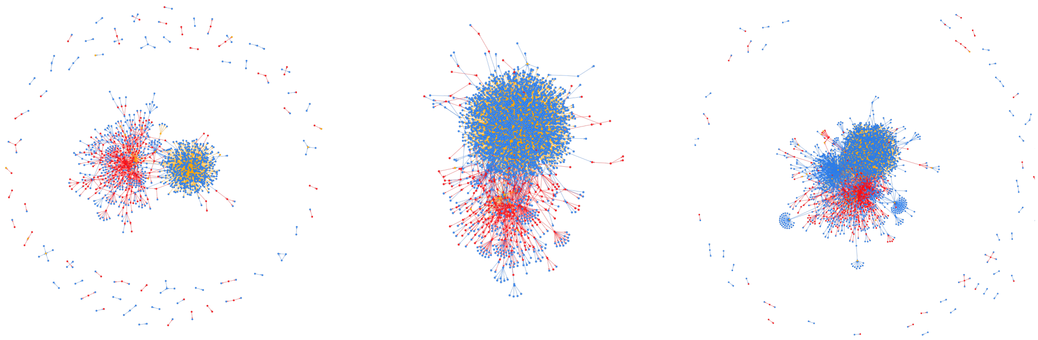
7.2 Future work

While our work is not the first to examine the EU ETS, it has provided so far the most complete picture of key network characteristics. This thesis though thorough, is by no means a comprehensive and complete study of every aspect of the network, and has room for expansion. While our testing of significant models in network theory is a start in the way of modeling and predicting network structure, there exist many models which might possibly provide a better fit to the static monthly networks. Another interesting area for further research is a model based on the temporal evolution of the network, that being the evolutionary modelling of its behavior. Although without abundant literature, temporal evolution of networks has been studied such as the work by Leifert et al [34] in which they examine TERGM(Temporal Exponential Random Graphs) and SAOM(Stochastic Actor-Oriented Model).With the EU ETS observed and analyzed as was done here, a more challenging step will be the modeling the behavior of the actors and use a game-theoretic perspective to approach it.

Chapter 8

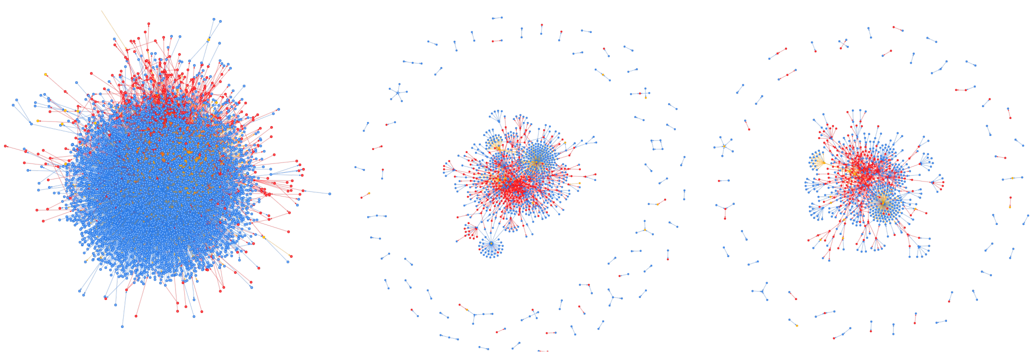
Appendix

In the following network snapshots, the regulated nodes are depicted in blue, the financial nodes are depicted in red and the governmental nodes are in yellow.



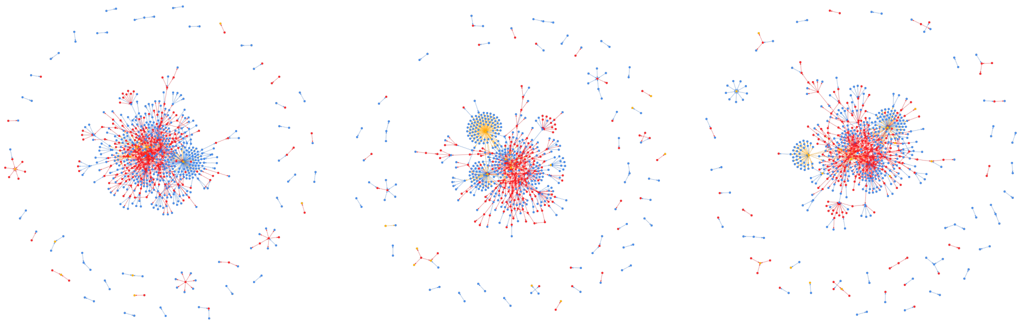
(a) January 2014 (b) February 2014 (c) March 2014

Figure 8.1: 1st quarter of 2014



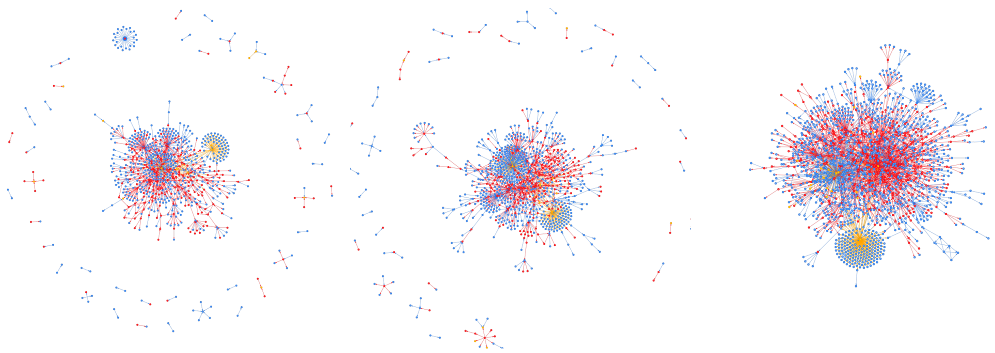
(a) April 2014 (b) May 2014 (c) June 2014

Figure 8.2: 2nd quarter of 2014



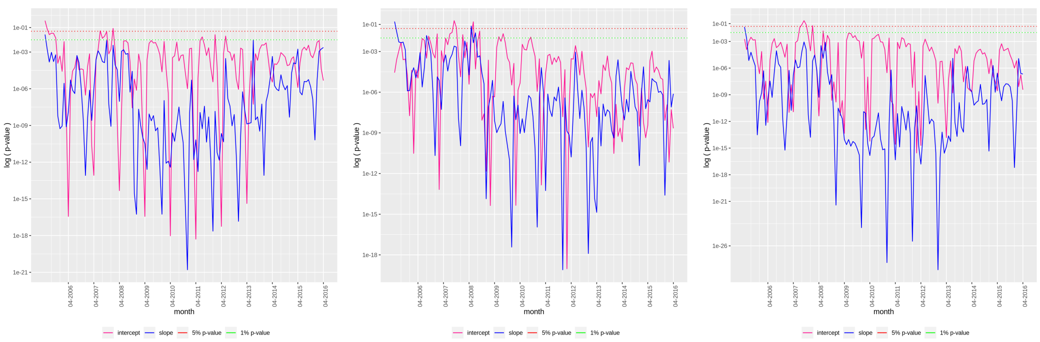
(a) July 2014 (b) August 2014 (c) September 2014

Figure 8.3: 3rd quarter of 2014



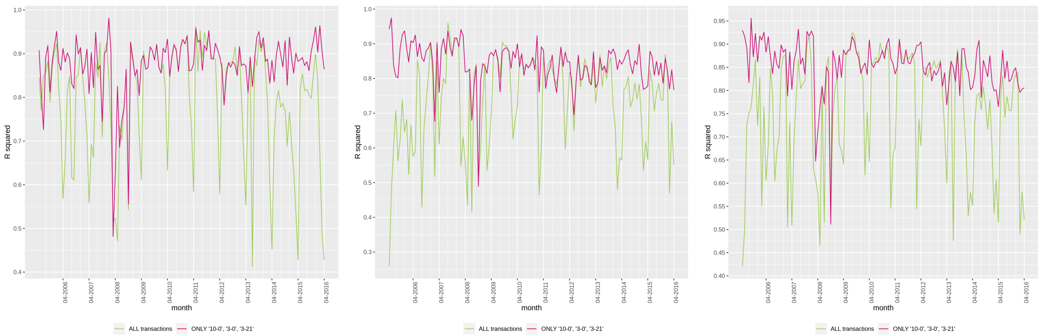
(a) October 2014 (b) November 2014 (c) December 2014

Figure 8.4: 4th quarter of 2014



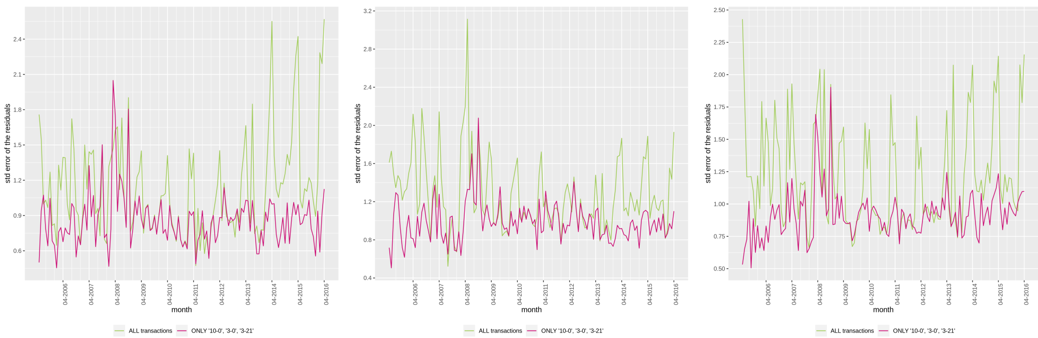
(a) In degree (b) Out degree (c) Total degree

Figure 8.5: p-values for Linear Regression



(a) In degree (b) Out degree (c) Total degree

Figure 8.6: R squared values of Linear Regression



(a) In degree (b) Out degree (c) Total degree

Figure 8.7: Standard Errors of Linear Regression

Bibliography

[1] A. Akerman and A. L. Seim. The global arms trade network 1950–2007. *Journal of Comparative Economics*, 42(3):535–551, 2014.

[2] I. Artico, I. Smolyarenko, V. Vinciotti, and E. Wit. How rare are power-law networks really? *Proceedings of the Royal Society A*, 476(2241):20190742, 2020.

[3] J. P. Bagrow and E. M. Bollt. An information-theoretic, all-scales approach to comparing networks. *Applied Network Science*, 4(1):1–15, 2019.

[4] J. P. Bagrow, E. M. Bollt, J. D. Skufca, and D. Ben-Avraham. Portraits of complex networks. *EPL (Europhysics Letters)*, 81(6):68004, 2008.

[5] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999.

[6] A.-L. Barabási. Network science. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 371(1987):20120375, Mar. 2013.

[7] S. Borghesi and A. Flori. Eu ets facets in the net: how account types influence the structure of the system. 2016.

[8] U. Brandes. On variants of shortest-path betweenness centrality and their generic computation. *Social Networks*, 30(2):136–145, 2008.

[9] A. D. Broido and A. Clauset. Scale-free networks are rare. *Nature communications*, 10(1):1–10, 2019.

[10] E. Burton and W. Sanjour. An economic analysis of the control of sulphur oxides air pollution. *DHEW Program Analysis Report. Ernst and Ernst. Washington, DC*, (69), 1967.

[11] E. Burton and W. Sanjour. Applications of cost-effectiveness analysis to air pollution control. *DHEW Contract No. CPA*, pages 22–69, 1970.

[12] A. Chandrasekhar. Econometrics of network formation. *The Oxford Handbook of the Economics of Networks*, pages 303–357, 2016.

[13] Council of European Union. Directive 2004/101/ec of the european parliament and the council, 2004.

[14] A. Creti, P.-A. Jouvet, and V. Mignon. Carbon price drivers: Phase i versus phase ii equilibrium? *Energy Economics*, 34(1):327–334, 2012.

[15] Á. De Paula. Econometric models of network formation. *Annual Review of Economics*, 12:775–799, 2020.

[16] B. Declercq, E. Delarue, and W. D'haeseleer. Impact of the economic recession on the european power sector's co2 emissions. *Energy Policy*, 39(3):1677–1686, 2011.

[17] S. Dimos, E. Evangelatou, D. Fotakis, A. Mantis, and A. Mathioudaki. Forecasting eu allowance price using transactions of central entities.

[18] S. Dimos, E. Evangelatou, D. Fotakis, A. Mantis, and A. Mathioudaki. On the impacts of allowance banking and the financial sector on the eu emissions trading system. *Euro-Mediterranean Journal for Environmental Integration*, 5(2):1–25, 2020.

[19] M. Falkenberg, J.-H. Lee, S.-i. Amano, K.-i. Ogawa, K. Yano, Y. Miyake, T. S. Evans, and K. Christensen. Identifying time dependence in network growth. *Physical Review Research*, 2(2):023352, 2020.

[20] D. N. Fisher, M. J. Silk, and D. W. Franks. The perceived assortativity of social networks: methodological problems and solutions. *Trends in Social Network Analysis*, pages 1–19, 2017.

[21] J. García-Algarra, M. L. Mouronte-López, and J. Galeano. A stochastic generative model of the world trade network. *Scientific reports*, 9(1):1–10, 2019.

[22] K. Goldner, N. Immorlica, and B. Lucier. Reducing inefficiency in carbon auctions with imperfect competition. *arXiv preprint arXiv:1912.06428*, 2019.

[23] E. M. Hafner-Burton, M. Kahler, and A. H. Montgomery. Network analysis for international relations. *International organization*, 63(3):559–592, 2009.

[24] P. Holme. Modern temporal network theory: a colloquium. *The European Physical Journal B*, 88(9):1–30, 2015.

[25] M. O. Jackson. A survey of models of network formation: Stability and efficiency. 2003.

[26] A. Karpf, A. Mandel, and S. Battiston. A network-based analysis of the european emission market. In *Proceedings of ECCS 2014*, pages 283–295. Springer, 2016.

[27] A. Karpf, A. Mandel, and S. Battiston. Price and network dynamics in the european carbon market. *Journal of Economic Behavior & Organization*, 153:103–122, 2018.

[28] J. J. Kay, L. A. Graham, and R. E. Ulanowicz. A detailed guide to network analysis. In *Network analysis in marine ecology*, pages 15–61. Springer, 1989.

[29] M.-K. Kim and T. Kim. Estimating impact of regional greenhouse gas initiative on coal to gas switching using synthetic control methods. *Energy Economics*, 59:328–335, 2016.

[30] D. Koutra, J. T. Vogelstein, and C. Faloutsos. Deltacon: A principled massive-graph similarity function. In *Proceedings of the 2013 SIAM International Conference on Data Mining*, pages 162–170. SIAM, 2013.

[31] P. Krapivsky and D. Krioukov. Scale-free networks as preasymptotic regimes of superlinear preferential attachment. *Physical Review E*, 78(2):026114, 2008.

[32] O. Kuchaiev, T. Milenković, V. Memišević, W. Hayes, and N. Pržulj. Topological network alignment uncovers biological function and phylogeny. *Journal of the Royal Society Interface*, 7(50):1341–1354, 2010.

[33] O. Kuchaiev and N. Pržulj. Integrative network alignment reveals large regions of global network similarity in yeast and human. *Bioinformatics*, 27(10):1390–1396, 2011.

[34] P. Leifeld and S. J. Cranmer. A theoretical and empirical comparison of the temporal exponential random graph model and the stochastic actor-oriented model. *Network Science*, 7(1):20–51, 2019.

[35] C. Leining, S. Kerr, and B. Bruce-Brand. The new zealand emissions trading scheme: Critical review and future outlook for three design innovations. *Climate Policy*, 20(2):246–264, 2020.

[36] W. W. Leontief. Quantitative input and output relations in the economic systems of the united states. *The Review of Economics and Statistics*, 18(3):105–125, 1936.

[37] J. Leskovec, J. Kleinberg, and C. Faloutsos. Graph evolution: Densification and shrinking diameters. *ACM transactions on Knowledge Discovery from Data (TKDD)*, 1(1):2–es, 2007.

[38] N. Malod-Dognin and N. Pržulj. L-graal: Lagrangian graphlet-based network aligner. *Bioinformatics*, 31(13):2182–2189, 2015.

[39] V. Memišević and N. Pržulj. C-graal: C ommon-neighbors-based global gra ph al ignment of biological networks. *Integrative Biology*, 4(7):734–743, 2012.

[40] T. Milenković, W. L. Ng, W. Hayes, and N. Pržulj. Optimal network alignment with graphlet degree vectors. *Cancer informatics*, 9:CIN–S4744, 2010.

[41] B. C. Murray and P. T. Maniloff. Why have greenhouse emissions in rggi states declined? an econometric attribution to economic, energy market, and policy factors. *Energy Economics*, 51:581–589, 2015.

[42] M. E. Newman. Mixing patterns in networks. *Physical review E*, 67(2):026126, 2003.

[43] M. E. Newman. The structure and function of complex networks. *SIAM review*, 45(2):167–256, 2003.

[44] Y. Nishida, Y. Hua, and N. Okamoto. Alternative building emission-reduction measure: outcomes from the tokyo cap-and-trade program. *Building Research & Information*, 44(5-6):644–659, 2016.

[45] R. Noldus and P. Van Mieghem. Assortativity in complex networks. *Journal of Complex Networks*, 3(4):507–542, 2015.

[46] M. R. Piraveenan. *Topological analysis of complex networks using assortativity*. University of Sydney, 2010.

[47] D. d. S. Price. A general theory of bibliometric and other cumulative advantage processes. *Journal of the American society for Information science*, 27(5):292–306, 1976.

[48] C. E. Shannon. A mathematical theory of communication. *ACM SIGMOBILE mobile computing and communications review*, 5(1):3–55, 2001.

[49] M. Tantardini, F. Ieva, L. Tajoli, and C. Piccardi. Comparing methods for comparing networks. *Scientific reports*, 9(1):1–19, 2019.

[50] UNFCCC. *Kyoto Protocol to the United Nations Framework*. 1997.

[51] E. Union. *EU Emissions Trading System (EU ETS)*.

[52] E. Union. *EU ETS Handbook*. 2015.

[53] J. Yan. The impact of climate policy on fossil fuel consumption: Evidence from the regional greenhouse gas initiative (rggi). *Energy Economics*, 100:105333, 2021.

[54] Ö. N. Yaveroğlu, N. Malod-Dognin, D. Davis, Z. Levnajic, V. Janjic, R. Karapandza, A. Stojmirovic, and N. Pržulj. Revealing the hidden language of complex networks. *Scientific reports*, 4(1):1–9, 2014.