A Note on 2–Degree Clustering and Dominating Ego Networks in a Undirected Graph

Moses Boudourides, *Moses.Boudourides@cs.haverford.edu*

Draft, January 2022

The theory of **structural holes**, originally developed by the seminal work of Ronald S. Burt (Burt, 1995), is based on combined occurrences of triadic (transitive) openings and closures in a graph, typically identified by examining the topological properties of 2–paths. Say $G = (V, E)$ is a undirected simple graph with set of nodes $V = \{1, 2, \ldots, n\}$ and set of edges $E = \{1, 2, \ldots, m\}$. Let "$\sim$" denote the adjacency relation on G, so that $i \sim j$ means that nodes i and j are adjacent (or neighbors). In this setting, a **2–path**, denoted as $P_2(i, j, k)$, is a sequence of three nodes $\{i, j, k\}$ such that $i \sim j$ and $j \sim k$; i and k are called **endpoints**, j **middle point**, and the 2–path is said to be **incident** to its endpoints. A 2–path $P_2(i, j, k)$ is called **open 2–path**, when its endpoints i and k are not adjacent, and **closed 2–path**, when they are. Apparently, an edge becomes a (**local**) **bridge**, when the endpoints of that edge do not share any other neighbors (Easley and Kleinberg, 2010)), meaning that, unless a bridge is an isolated pair of adjacent nodes, any (other) bridge should necessarily show up at one of the two edges contained in an open 2–path. Furthermore, a closed 2–path always forms a 3–cycle (**triangle**). As a matter of fact, Duncan Watts and Steven Strogatz's measure of (local) **clustering coefficient** of a node j (Watts and Strogatz, 1998), can be defined through 2–paths as follows:

$$C(j) = \frac{\tau(j)}{\pi_2(j)},$$

where $\tau(j)$ and $\pi_2(j)$ denote the number of all closed or open (respectively) 2–paths both with middle point j. Obviously, $\tau(j)$ is the number of all triangles having j as one of their vertices and $\pi_2(j)$ is the number of all pairs of neighbors of j, i.e., $\pi_2(j) = \binom{d_j}{2} = \frac{1}{2}d_j(d_j - 1)$, where d_j is the **degree** of node j. In *Structural Holes* (Burt, 1995), Burt has given an equivalent measure of (local) clustering, called **redundancy**, based on ego (sub)networks. Given a node j, the **egocentric subgraph** (or **ego network**) around j is the induced subgraph of j's (**open**) **neighborhood set** $N(j)$, composed of j (called **ego**) and all j's neighbors (called **alters**) (Perry et al., 2018). As Stephen Borgatti (Borgatti, 1997) has shown, (nonisolate) node j's **redundancy** is given by

$$\text{redundancy}(j) = \frac{2\tau(j)}{d_j} = (d_j - 1)C(j).$$

Through the formalism of 2–paths, one could build an equivalent notion to that of the clustering coefficient using the nodal 2–degree, which can be defined in one of the following two ways:

- As the **total neighborhood degree**, denoted by t_j, where t_j is defined to be the sum of degrees of all neighbors k of j, i.e., $t_j = \sum_{k \sim j} d_k$ (Cao, 1998).

- As the **2–path degree**, denoted by d_{2j}, where d_{2j} is defined to be the total number of 2–paths incident to j, i.e., symbolically, $d_{2j} = |P_2(j)|$, where $P_2(j)$ is the set of all 2–paths with endpoint j. Notice that, for any node j,

$$d_{2j} = t_j - d_j.$$

Recall that a $n \times m$ matrix B is the (unoriented) **incidence matrix** of the graph G if $B_{ij} = 1$, whenever node i and edge j are incident, while $B_{ij} = 0$, otherwise. Thus,

$$BB^T = D + A,$$

where exponent T designates transpose, D is the $n \times n$ diagonal matrix with the degrees of nodes of G along the diagonal, and A is the adjacency matrix of G, i.e., a $n \times n$ matrix such that $A_{ij} = 1$, whenever $i \sim j$, and $A_{ij} = 0$, otherwise.

Similarly, denoting by m_2 the total number of 2–paths in graph G:

$$m_2 = \sum_{j=1}^{n} \binom{d_j}{2} = \frac{1}{2} \sum_{j=1}^{n} d_j(d_j - 1),$$

there exists a $n \times m_2$ matrix B_2, called (unoriented) **2–incidence matrix** of G, such that $B_{2ij} = 1$, whenever node i and 2–path j are incident, while $B_{2ij} = 0$, otherwise. It is not difficult to show that:

$$B_2 B_2^T = D_2 + A^2,$$

where D_2 is the $n \times n$ diagonal matrix with the 2–degrees of nodes of G along the diagonal. In addition, for any node $i \in V$, one might get the nodal degrees and 2–degrees from the incidence or the adjacency matrix of the graph as follows:

$$d_i = \sum_{j=1}^{m} B_{ij} = \sum_{j=1}^{n} A_{ij},$$
$$d_{2i} = \sum_{j=1}^{m_2} B_{2ij} = \sum_{j=1}^{n} (A_{ij} - 1) d_j.$$

Furthermore, the **handshaking lemmas** for the degree and the 2–degree are:

$$\sum_{i=1}^{n} d_i = 2m,$$
$$\sum_{i=1}^{n} d_{2i} = 2m_2.$$

Now, since a 2–path can be either closed or open, for any node i, the 2–degree d_{2i} can be decomposed to a **2–degree pair** $(d_{2_i}^{\text{clustered}}, d_{2_i}^{\text{traversing}})$, where $d_{2_i}^{\text{clustered}}$ is the **clustered–2–degree** of i, which is equal to the number of all closed 2–paths incident to i, i.e., $d_{2_i}^{\text{clustered}} = \tau(i)$, and $d_{2_i}^{\text{traversing}}$ is the **traversing–2–degree** of i, which is equal to the number of all open 2–paths incident to i, where

$$d_{2_i}^{\text{clustered}} + d_{2_i}^{\text{traversing}} = d_{2i}.$$

Of course, denoting by τ the total number of triangles in graph G ($\tau = \frac{1}{6}\operatorname{Tr}(A^3)$), there exists a $n \times \tau$ matrix $B_2^{\text{clustered}}$, called (unoriented) **clustered–2–incidence matrix** of G (defined as:

${B_2}_{ij}^{\text{clustered}} = 1$, whenever i is one of the vertices of triangle j, ${B_2}_{ij}^{\text{clustered}} = 0$, otherwise), such that, for any (nonisolate) node i:

$$\begin{aligned} {d_2}_i^{\text{clustered}} &= \sum_{j=1}^{\tau} {B_2}_{ij}^{\text{clustered}} = \frac{1}{2}\{A^3\}_{ii} = \tau(i) = \\ &= \frac{1}{2}\, d_i \,\text{redundancy}(i) = \frac{1}{2}\, d_i(d_i - 1)C(i). \end{aligned}$$

In any case, in a graph (having diameter larger or equal to 2), the clustering coefficient (or redundancy or clustered–2–degree) of every node j is either positive (i.e., j is a vertex in at least one triangle) or it is zero (i.e., the only edges incident to j are bridges). In the former case, let us call j **clustered node** and denote by $V_{\text{clustered}}$ the set of all clustered nodes. In the latter case, let us call j **traversing node** and denote by $V_{\text{traversing}}$ the set of all traversing nodes. One can easily show the following statements:

1. If G is complete, all nodes are clustered (i.e., $V_{\text{clustered}} = V$ and $|V_{\text{traversing}}| = 0$).
2. If G is either a tree or a k–cycle graph (where $k > 3$) or a bipartite graph or a graph without triangles, all nodes are traversing (i.e., $V_{\text{traversing}} = V$ and $|V_{\text{clustered}}| = 0$).
3. A graph G contains triangles (i.e., it has $\tau > 0$) if and only if $V_{\text{clustered}} \neq \varnothing$, in which case $3 \leq |V_{\text{clustered}}| \leq 3\tau$.

Correspondingly, let us call **nontrivial ego network** an egocentric subgraph focused around a clustered node and **bridging ego network** an egocentric subgraph focused around a traversing node.

Definition. Let G be a graph with $V_{\text{clustered}} \neq \varnothing$ and let $G(V_{\text{clustered}})$ be the subgraph in G of the set of all clustered nodes. If S is the *minimum dominating set* of $G(V_{\text{clustered}})$, then, for any $s \in S$, the nontrivial ego network centered at s is called **dominating nontrivial ego network** in G. Additionally, for any $t \in V_{\text{clustered}} \setminus S$, the nontrivial ego network centered at t is called **dominated nontrivial ego network** in G.

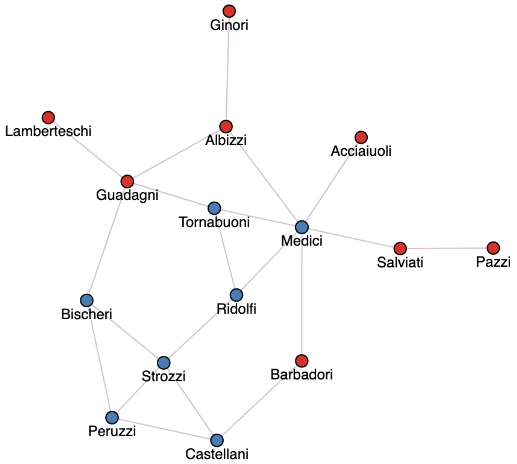
Notice that, by definition, $V_{\text{clustered}}$ breaks into two sets of nodes: the set of dominating and the set of dominated clustered nodes. On the other side, $V_{\text{traversing}}$ can break into the following four sets of traversing nodes (some of which it is possible that they might be empty):

- T_1, is the set of traversing nodes, which are adjacent to at least one dominating ego in the set S, but not adjacent to any dominated ego in $V_{\text{clustered}} \setminus S$.
- T_2, is the set of traversing nodes, which are adjacent to both a dominating ego in the set S and a dominated ego in $V_{\text{clustered}} \setminus S$.
- T_3, is the set of traversing nodes, which are adjacent to at least one dominated ego in the set $V_{\text{clustered}} \setminus S$, but not adjacent to any dominating ego in S.
- T_4, is the set of traversing nodes, which are not adjacent to any clustered node.

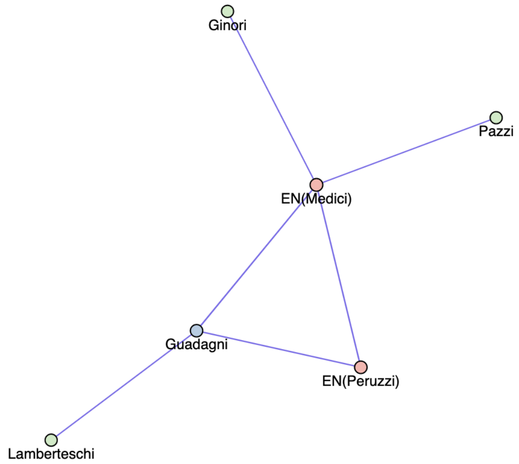
In order to grasp the complex ways through which dominating nontrivial ego networks are interwined with traversing nodes (which, of course, form bridging ego networks, i.e., trivial ego networks, since their clustering is zero), one might want to derive a composite (heterogeneous) graph with nodes of two types: dominating nontrivial ego networks and possibly traversing nodes of the type T_3 or T_4. Let us call this graph "**coupled graph of ego networks and traversing nodes**." Apparently, this is a multigraph, because by aggregating nodes into the dominating nontrivial egonets, one should also aggregate simple into parallel edges between or among ego networks and traversing nodes. In the sequel, we are displaying the plots of ten (small) graphs together with the corresponding coupled graphs and also their degree and 2–degree distributions. (Interactive plots of these fraphs are available at `http://mboudour.github.io/2022/01/06/Graphs-With-Triadic-Closure-Openness.html`.)

The Florentine families graph

Graph of 7 clustered and 8 traversing nodes

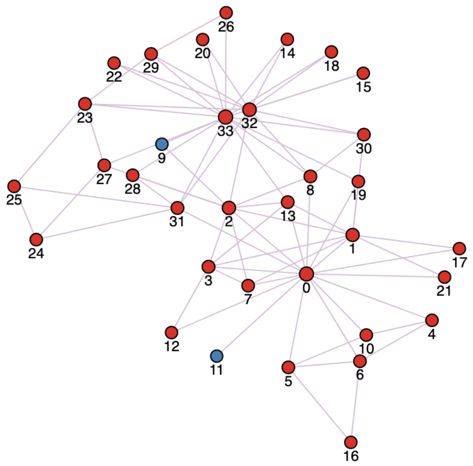


Coupled graph of 2 dominating ego networks with 4 traversing nodes

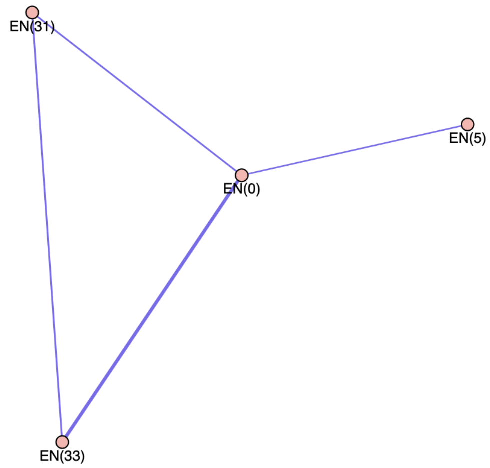


The Karate club graph

Graph of 32 clustered and 2 traversing nodes

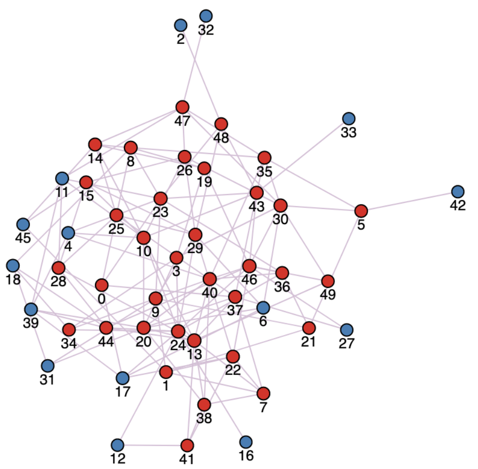


Coupled graph of 4 dominating ego networks with 0 traversing nodes

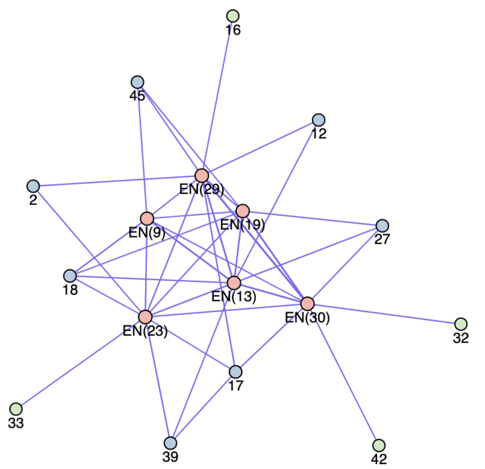


Erdos-Renyi random graph with n=50 and p=0.10

Graph of 35 clustered and 15 traversing nodes

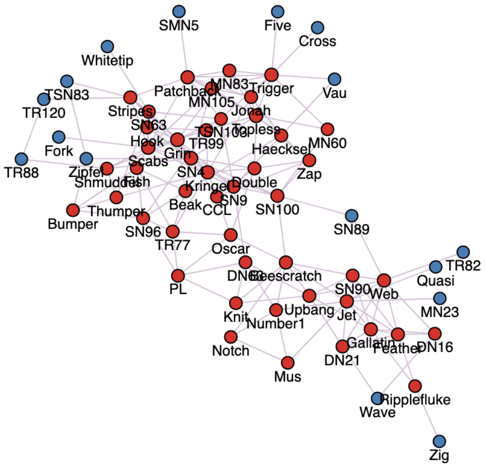


Coupled graph of 6 dominating ego networks with 11 traversing nodes

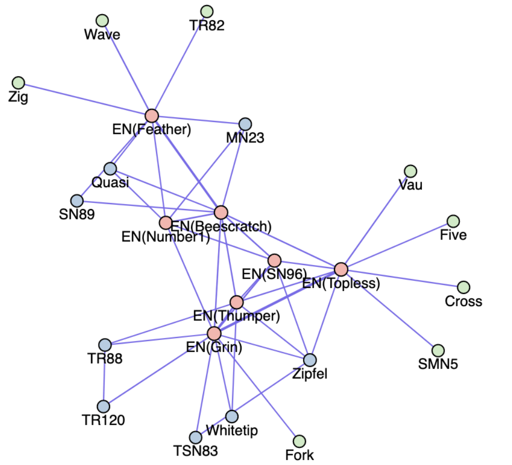


Newman's dolphins graph

Graph of 46 clustered and 16 traversing nodes

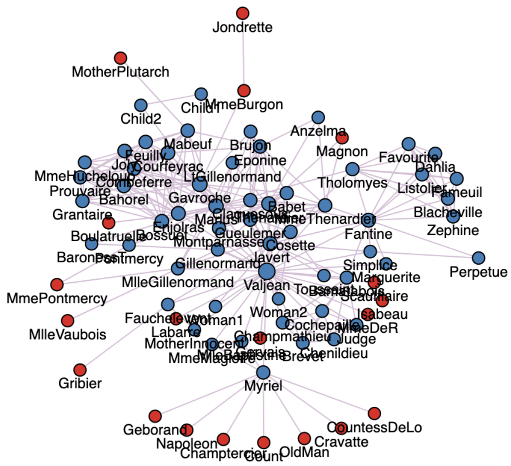


Coupled graph of 7 dominating ego networks with 16 traversing nodes

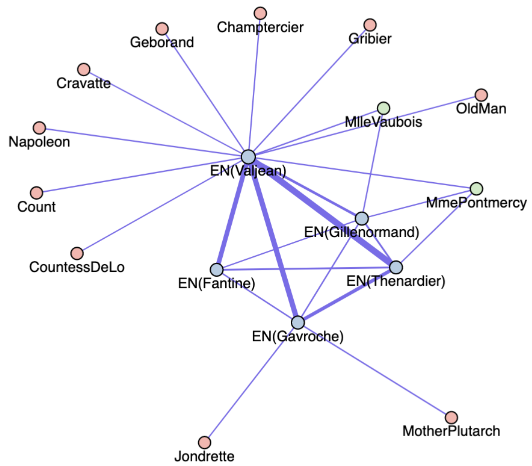


Les Miserables graph

Graph of 57 clustered and 20 traversing nodes

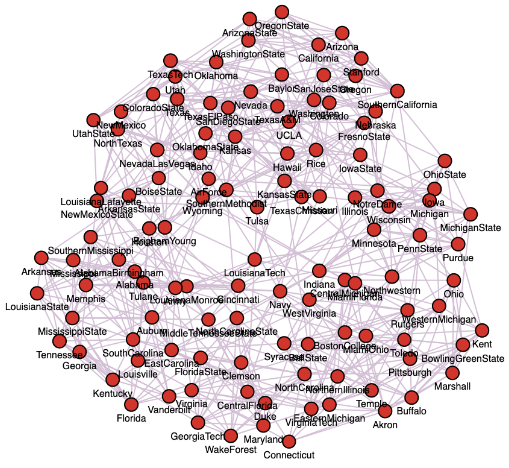


Coupled graph of 5 dominating ego networks with 12 traversing nodes

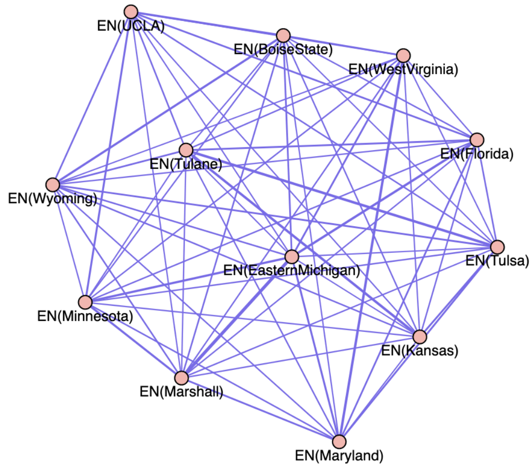


US college football graph

Graph of 115 clustered and 0 traversing nodes

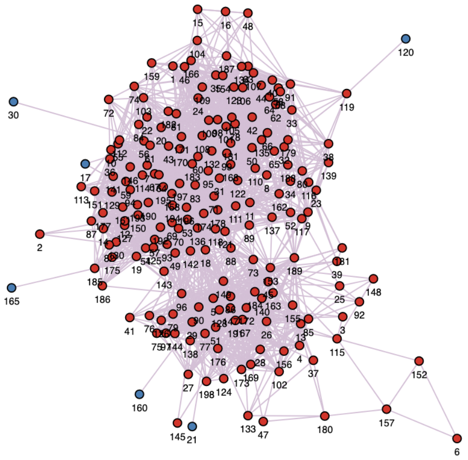


Coupled graph of 12 dominating ego networks with 0 traversing nodes

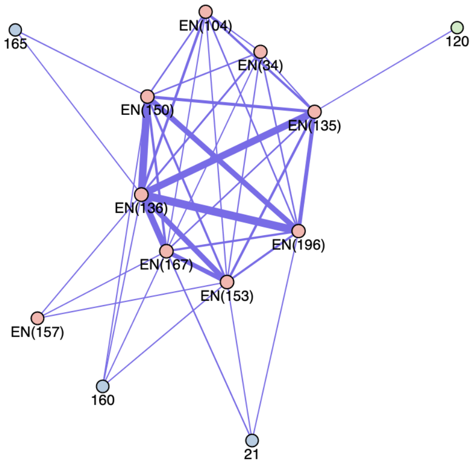


Jazz musicians graph

Graph of 192 clustered and 6 traversing nodes

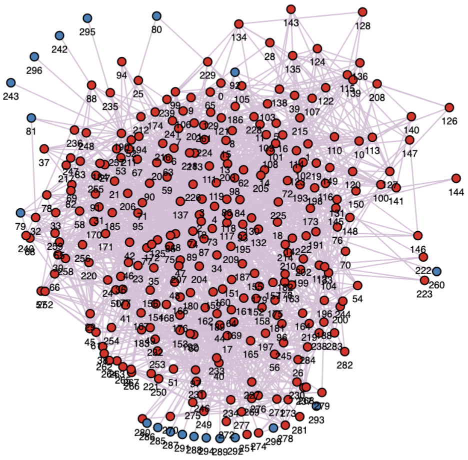


Coupled graph of 9 dominating ego networks with 4 traversing nodes

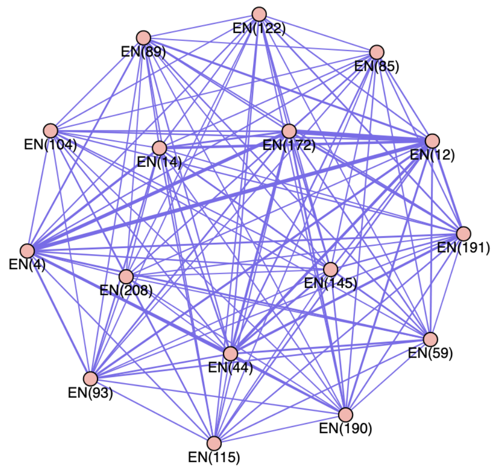


C. elegans graph

Graph of 278 clustered and 19 traversing nodes

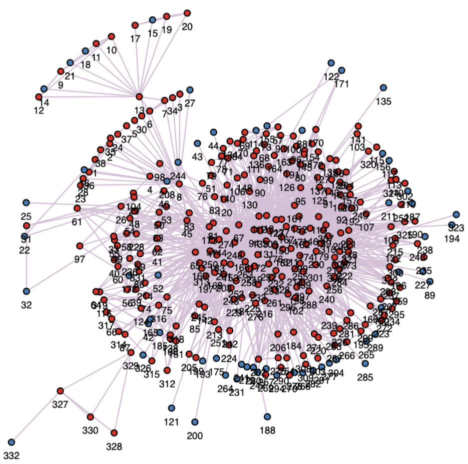


Coupled graph of 16 dominating ego networks with 0 traversing nodes

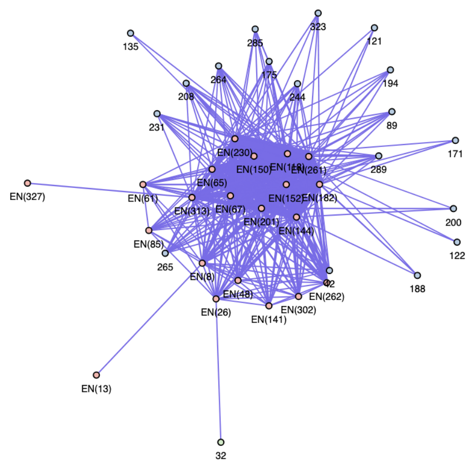


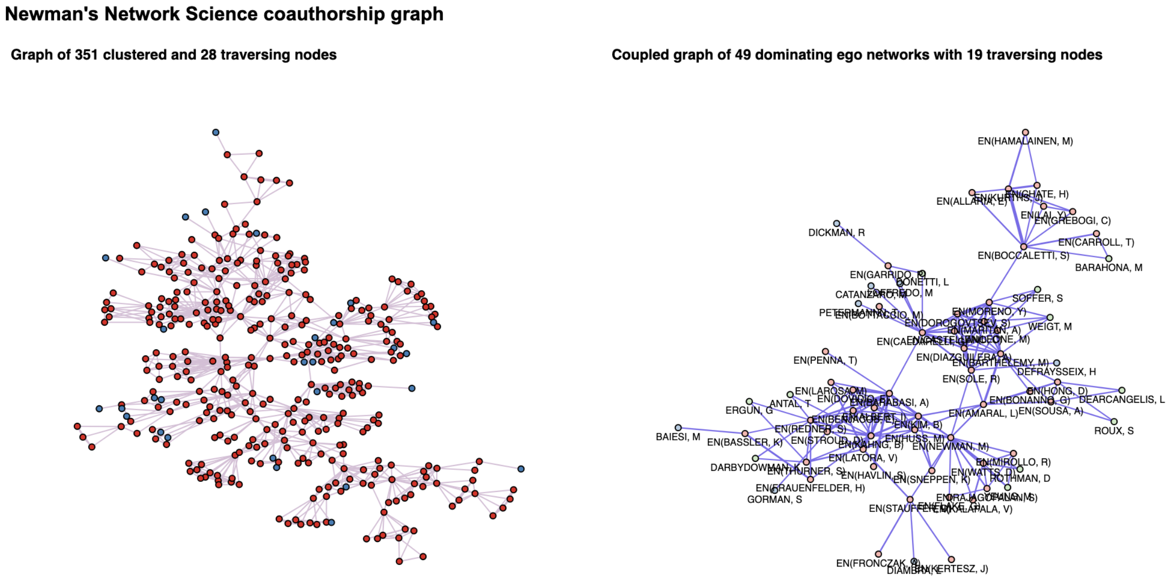
US air transportation graph

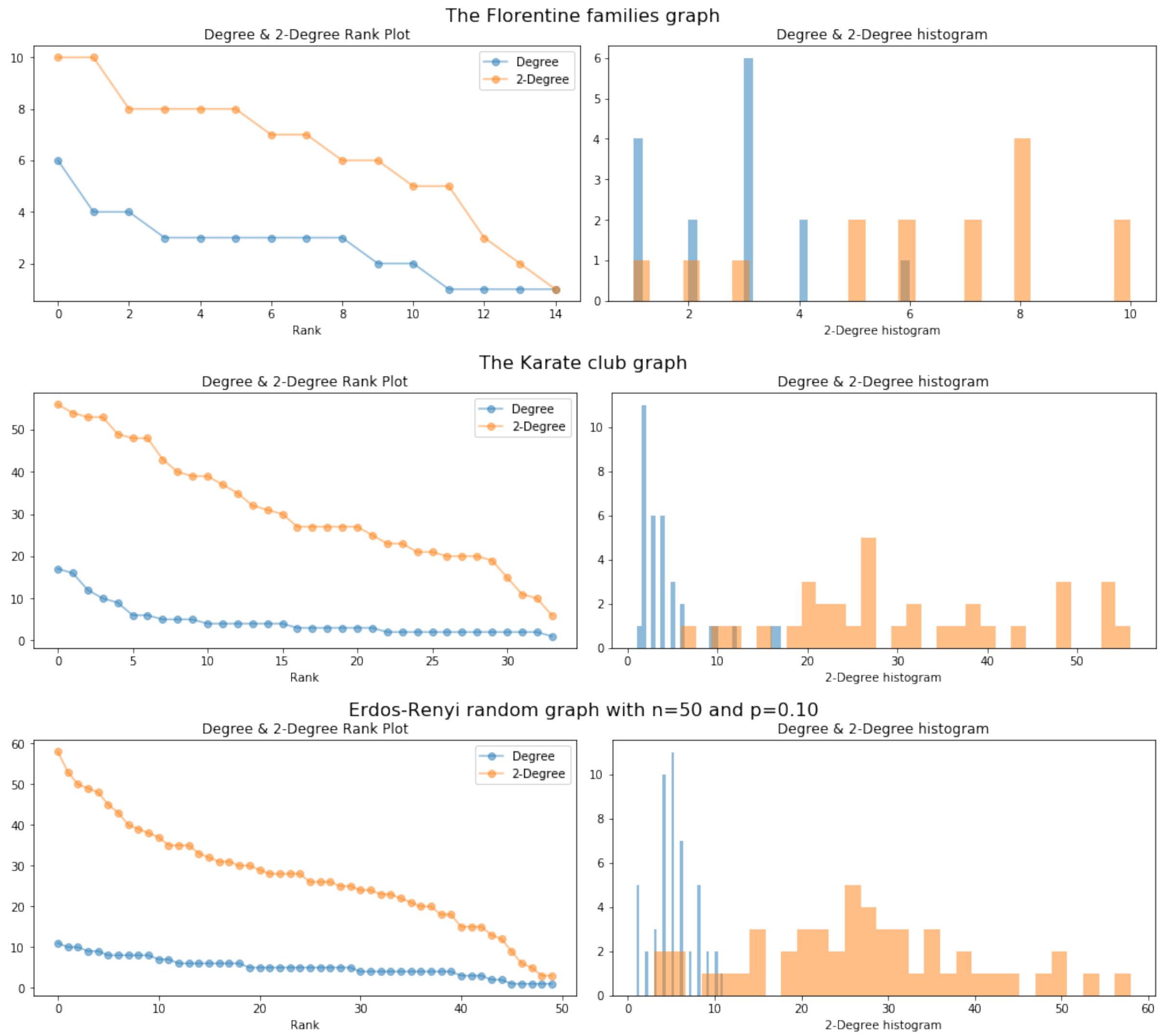
Graph of 272 clustered and 60 traversing nodes

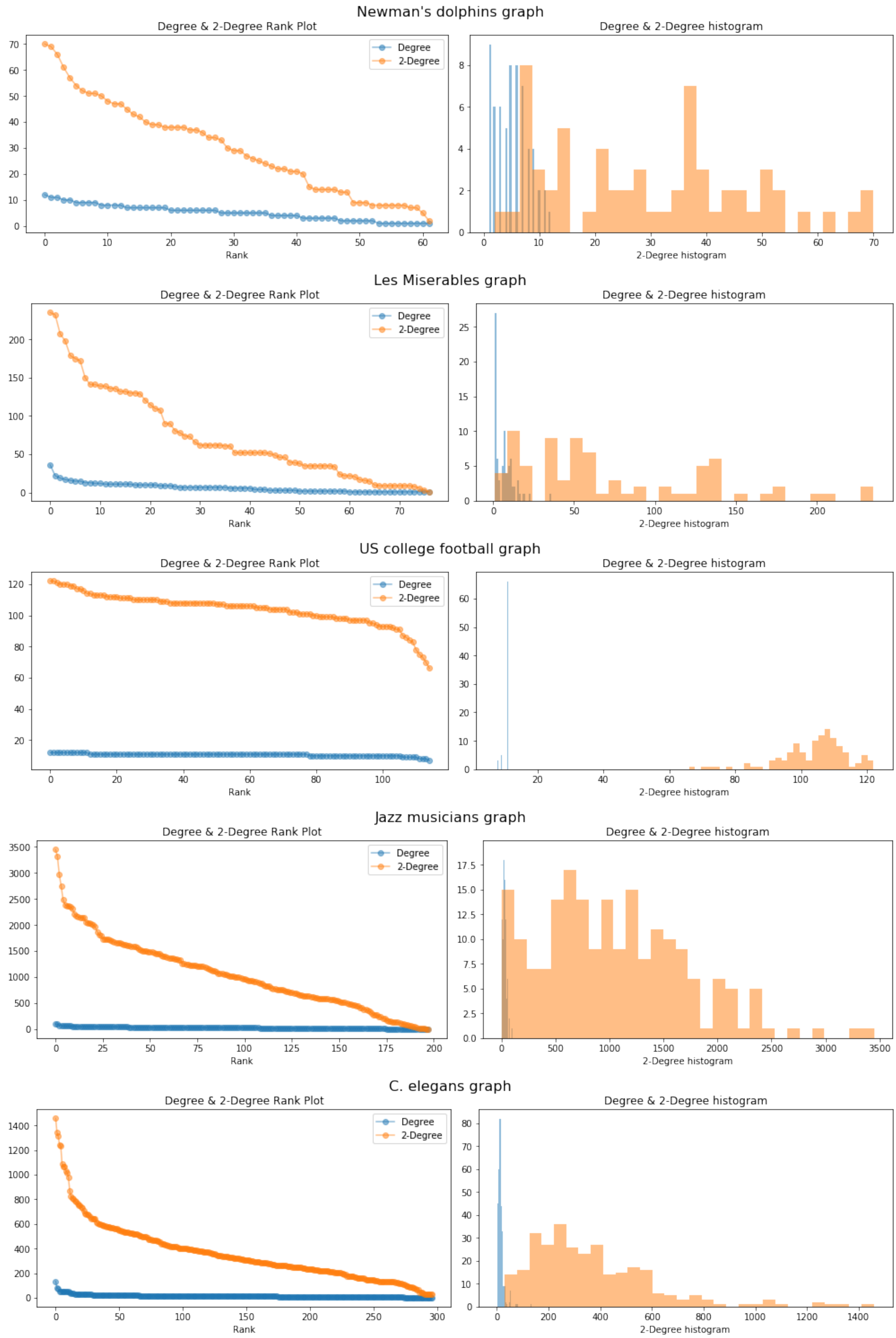


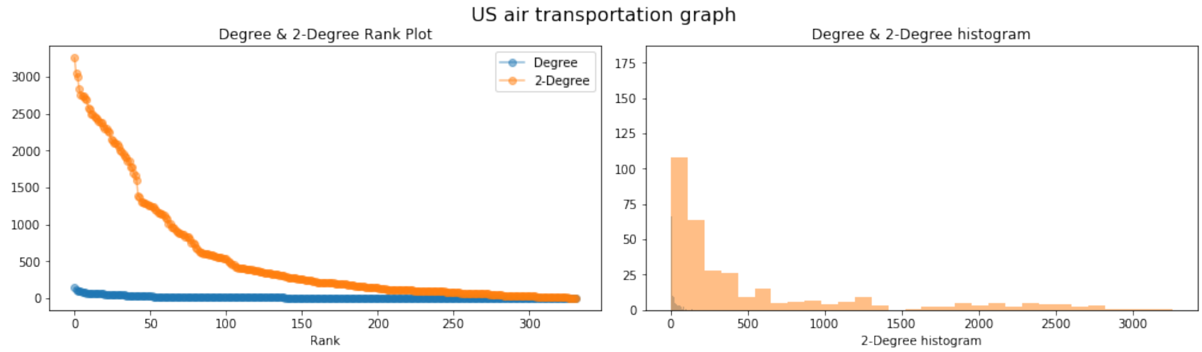
Coupled graph of 21 dominating ego networks with 19 traversing nodes

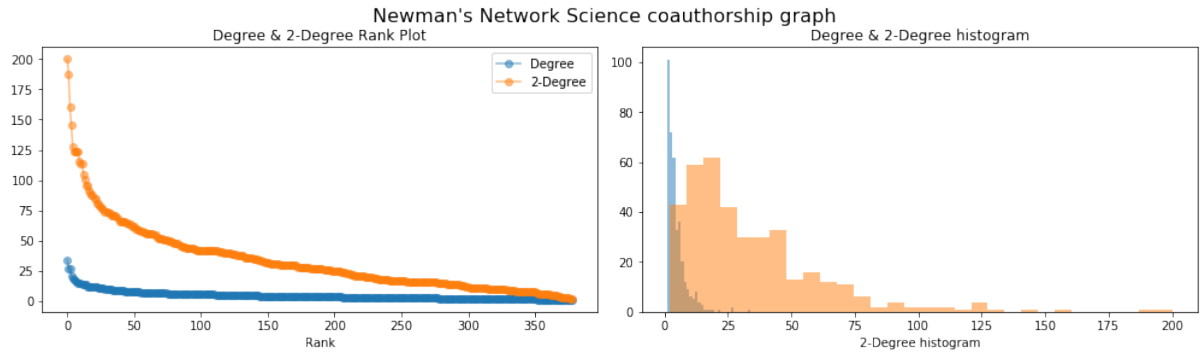


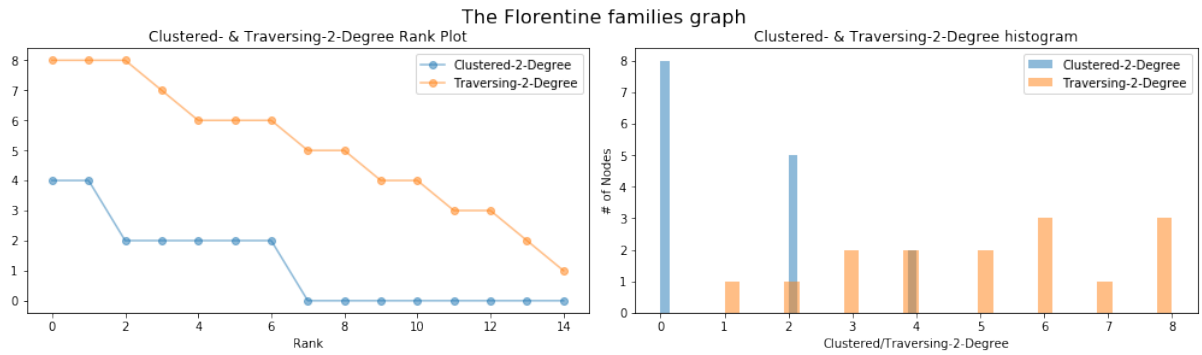

Newman's Network Science coauthorship graph
Graph of 351 clustered and 28 traversing nodes
Coupled graph of 49 dominating ego networks with 19 traversing nodes

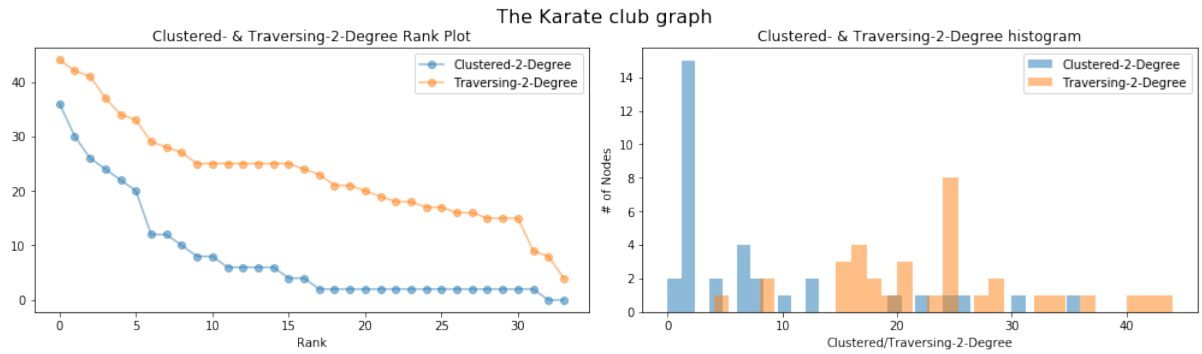

The Florentine families graph
Degree & 2-Degree Rank Plot
Degree
2-Degree
Rank
Degree & 2-Degree histogram
2-Degree histogram
The Karate club graph
Degree & 2-Degree Rank Plot
Degree
2-Degree
Rank
Degree & 2-Degree histogram
2-Degree histogram
Erdos-Renyi random graph with n=50 and p=0.10
Degree & 2-Degree Rank Plot
Degree
2-Degree
Rank
Degree & 2-Degree histogram
2-Degree histogram

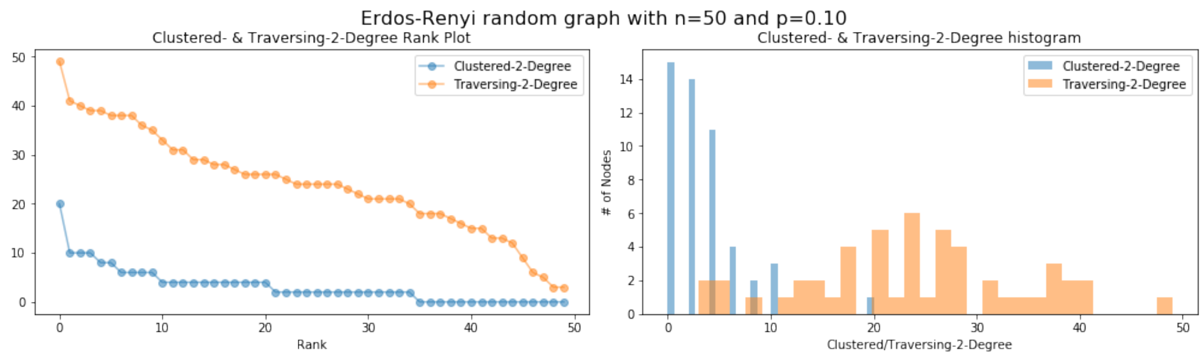

Newman's dolphins graph
Degree & 2-Degree Rank Plot
Degree
2-Degree
Rank
Degree & 2-Degree histogram
2-Degree histogram
Les Miserables graph
Degree & 2-Degree Rank Plot
Degree
2-Degree
Rank
Degree & 2-Degree histogram
2-Degree histogram
US college football graph
Degree & 2-Degree Rank Plot
Degree
2-Degree
Rank
Degree & 2-Degree histogram
2-Degree histogram
Jazz musicians graph
Degree & 2-Degree Rank Plot
Degree
2-Degree
Rank
Degree & 2-Degree histogram
2-Degree histogram
C. elegans graph
Degree & 2-Degree Rank Plot
Degree
2-Degree
Rank
Degree & 2-Degree histogram
2-Degree histogram


US air transportation graph
Degree & 2-Degree Rank Plot
Degree
2-Degree
Rank
Degree & 2-Degree histogram
2-Degree histogram

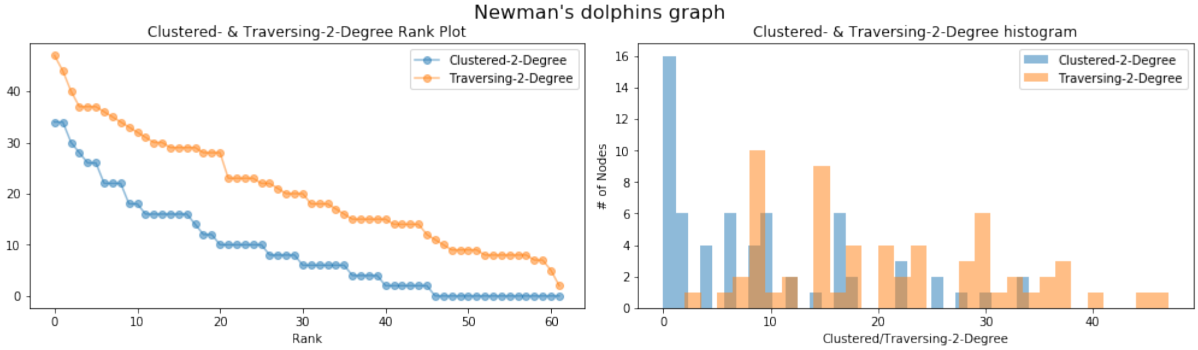

Newman's Network Science coauthorship graph
Degree & 2-Degree Rank Plot
Degree
2-Degree
Rank
Degree & 2-Degree histogram
2-Degree histogram


The Florentine families graph
Clustered- & Traversing-2-Degree Rank Plot
Clustered-2-Degree
Traversing-2-Degree
Rank
Clustered- & Traversing-2-Degree histogram
# of Nodes
Clustered/Traversing-2-Degree

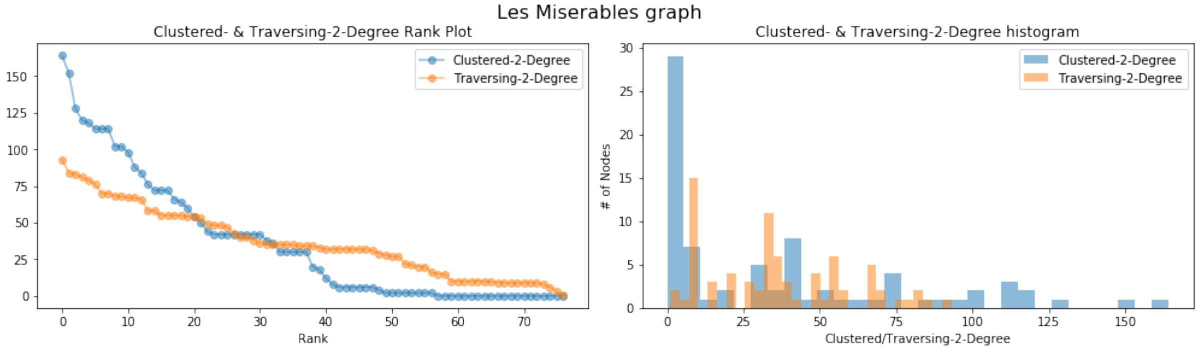

The Karate club graph
Clustered- & Traversing-2-Degree Rank Plot
Clustered-2-Degree
Traversing-2-Degree
Rank
Clustered- & Traversing-2-Degree histogram
# of Nodes
Clustered/Traversing-2-Degree


Erdos-Renyi random graph with n=50 and p=0.10
Clustered- & Traversing-2-Degree Rank Plot
Clustered-2-Degree
Traversing-2-Degree
Rank
Clustered- & Traversing-2-Degree histogram
# of Nodes
Clustered/Traversing-2-Degree

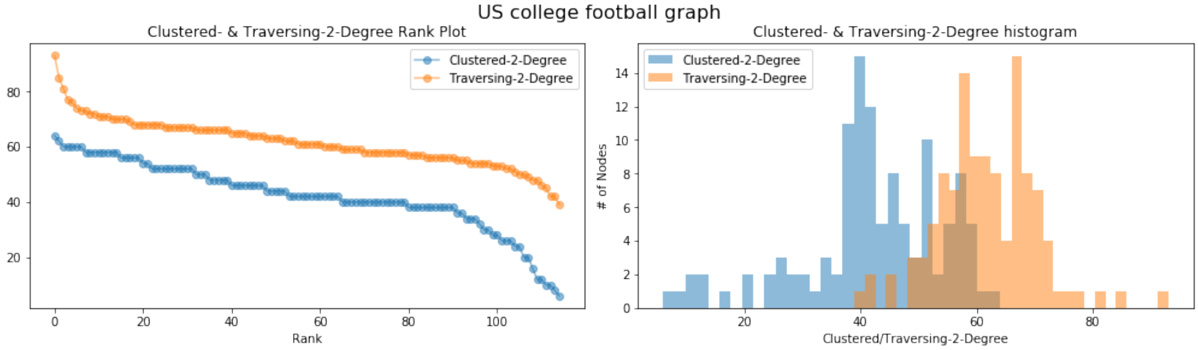
Newman's dolphins graph



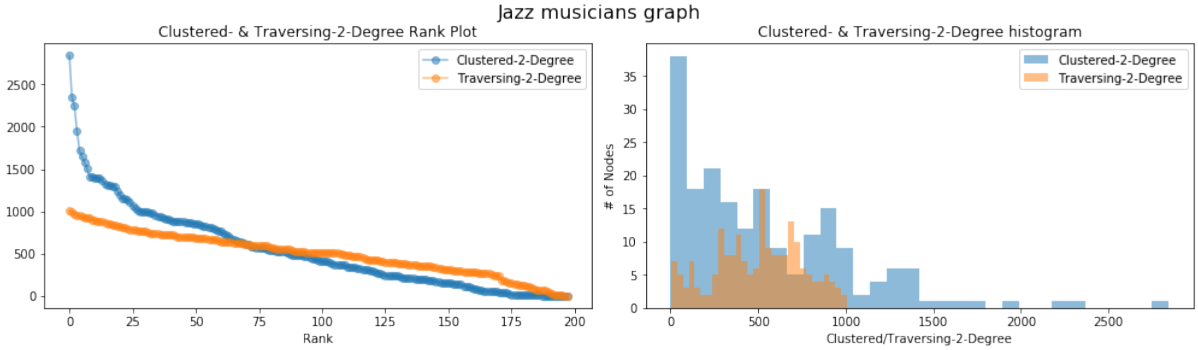
Les Miserables graph



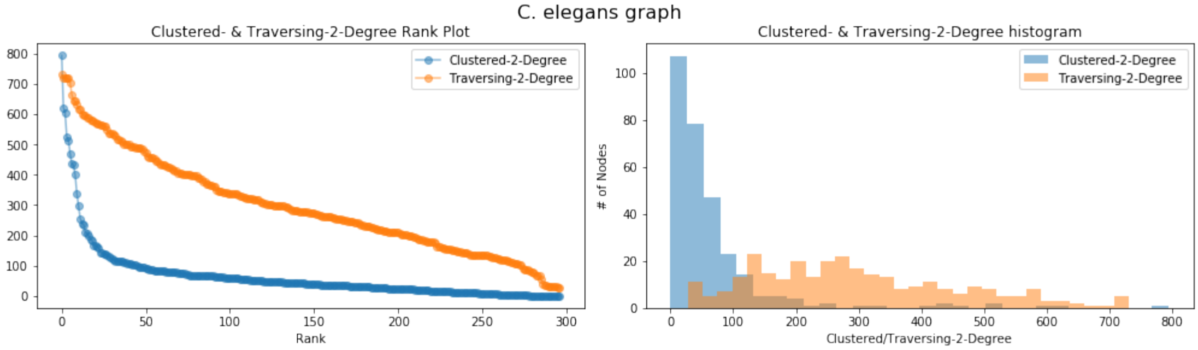
US college football graph

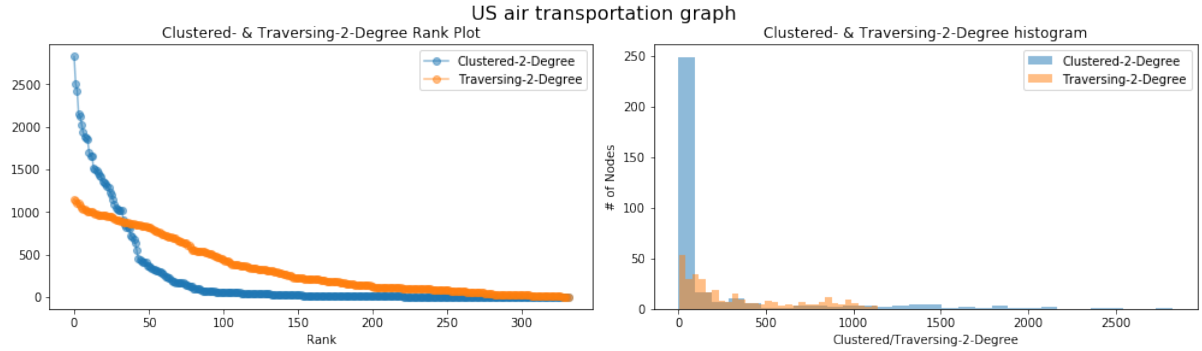


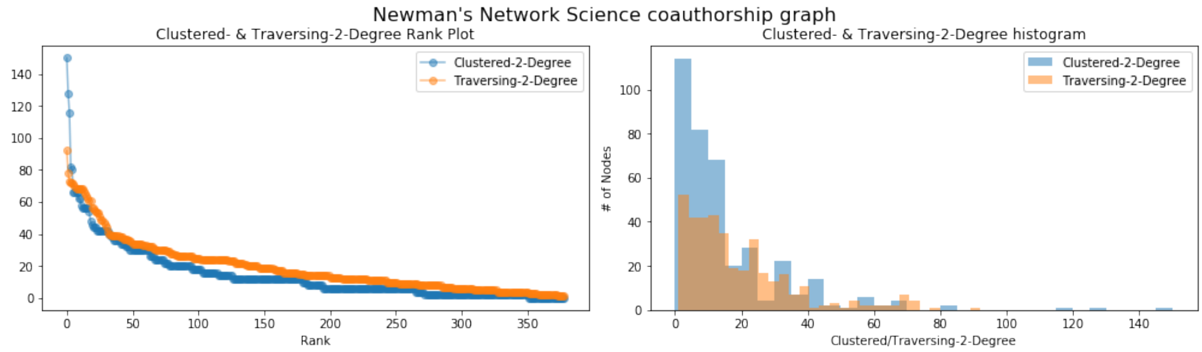
Jazz musicians graph



C. elegans graph







References

Borgatti, Stephen (1997). Structural Holes: Unpacking Burt's Redundancy Measures. *Connections*. INSNA.

Burt, Ronald S. (1995). *Structural Holes: The Social Structure of Competition*. Cambridge: Harvard University Press.

Cao, Dasong (1998). Bounds on eigenvalues and chromatic numbers. *Linear Algebra Appl.*, vol. 270, pp. 1–13.

Easley, David, and Kleinberg, Jon (2010). *Networks, Crowds, and Markets: Reasoning about a Highly Connected World*. Cambridge, UK: Cambridge University Press.

Perry, Brea L., Pescosolido, Bernice A., and Borgatti, Stephen P. (2018). *Egocentric Network Analysis: Foundations, Methods, and Models*. Cambridge: Cambridge University Press.

Watts, Duncan J., and Strogatz, Steven H. (1998). Collective dynamics of "small–world" networks. *Nature*, vol. 393, pp. 440–442.