Stochastic Network Models

Des Higham
School of Mathematics
University of Edinburgh

Motivation

Large-scale data generated by, or relevant to, human behaviour, e.g., **social media**, **on-line behaviour**

Potential to

- validate theories from **social science**
- inform **customer-facing** industries and organisations

Based on material from
Centrality-friendship paradoxes: When our friends are more important than us , D. J. Higham, Journal of Complex Networks, 2018
Infering and Calibrating Triadic Closure in a Dynamic Network, A. V. Mantzaris and D. J. Higham, in Temporal Networks, edited by P. Holme and J. Saramaki, 2013
Bistability through triadic closure, P. Grindrod, D. J. Higham and M. C. Parsons, Internet Mathematics, 2012
Models for evolving networks: with applications in telecommunication and online activities, P. Grindrod and D. J. Higham, IMA J. Management Mathematics, 2012

Part 1

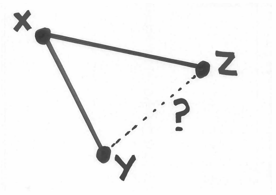
Triadic Closure. . .

Triadic Closure

Suggested by German sociologist **Georg Simmel** in 1908

Popularized by US sociologist **Mark Granovetter** in 1973

In terms of friendships, suppose X is a friend of Y, and X is a friend of Z, but Y is not a friend of Z



Then **Y is likely to become friends with Z**

Reasons include:

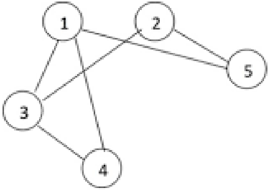
- Y is likely to meet Z
- Y and Z are vouched for by X
- X saves time/energy if Y and Z become friends

Simple Unweighted Graph

Adjacency matrix A Graph G

0	0	1	1	1
0	0	1	0	1
1	1	0	1	0
1	0	1	0	0
1	1	0	0	0

(a)



(b)

For $i \neq j$, the expression

$$\left(A^2\right)_{ij} := \sum_{p=1}^{n} a_{ip} a_{pj}$$

counts the **number of friends that nodes i and j have in common**

Develop a time-dependent model…

Evolving Graph Model Framework

Fixed number of of nodes, n

Evolving Graph Model Framework

Fixed number of of nodes, n

Edges may appear or disappear at discrete timepoints $0, 1, 2, 3, \ldots$

Evolving Graph Model Framework

Fixed number of of nodes, n

Edges may appear or disappear at discrete timepoints $0, 1, 2, 3, \ldots$

At each timepoint we have an undirected graph, represented by a symmetric, binary adjacency matrix $A^{[k]}$

Evolving Graph Model Framework

Fixed number of of nodes, n

Edges may appear or disappear at discrete timepoints $0, 1, 2, 3, \ldots$

At each timepoint we have an undirected graph, represented by a symmetric, binary adjacency matrix $A^{[k]}$

Our model will be a **discrete time Markov chain**. Given $A^{[k]}$ at time k there is a transition probability to every possible adjacency matrix at time $k+1$

Evolving Graph Model Framework

Fixed number of of nodes, n

Edges may appear or disappear at discrete timepoints $0, 1, 2, 3, \dots$

At each timepoint we have an undirected graph, represented by a symmetric, binary adjacency matrix $A^{[k]}$

Our model will be a **discrete time Markov chain**. Given $A^{[k]}$ at time k there is a transition probability to every possible adjacency matrix at time $k + 1$

To simplify the framework, we assume **edge independence**: between timepoints, the probability of an edge appearing or disappearing is independent of that for all other edges

Triadic Closure Model

Friends of friends become friends

We have n people, "friending" and "unfriending"
$A^{[k]}$ is the adjacency matrix at time k

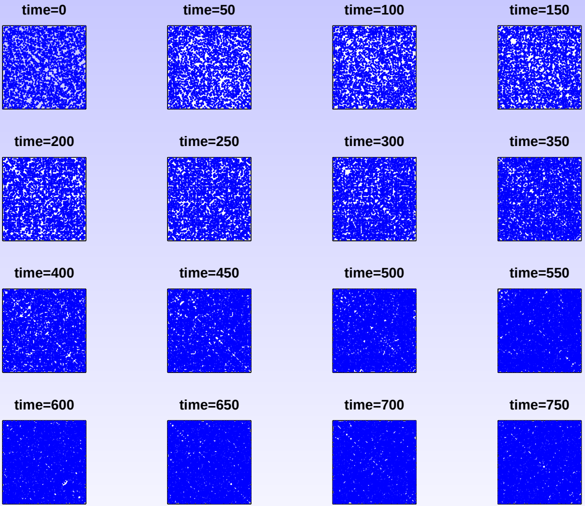
Edge death probability is a constant $\omega \in (0, 1)$
Edge birth probability between nodes i and j given by

$$\delta + \epsilon \left(\left(A^{[k]} \right)^2 \right)_{ij}$$

where $0 < \delta \ll 1$ and $0 < \epsilon(n-2) < 1 - \delta$

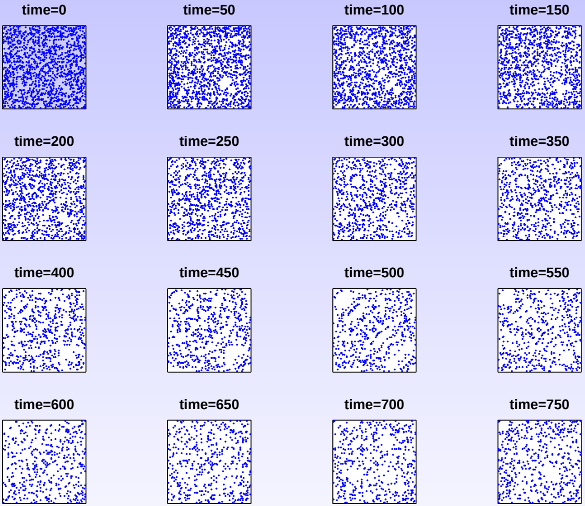
Consider $n = 100$, $\omega = 0.01$, $\epsilon = 5 \times 10^{-4}$, $\delta = 4 \times 10^{-4}$

Triadic closure: start with ER(0.3)



Edge density at time 750 is 0.712

Triadic closure: start with ER(0.15)



Edge density at time 750 is 0.051

Mean field analysis for $\delta + \epsilon \left(\left(A^{[k]}\right)^2 \right)_{ij}$

Ergodicity and **symmetry** $\Rightarrow$ Erdös-Rényi limit: every edge present with probablity $p^\star$

Heuristic **mean field** approach: insert the ansatz "$A^{[k]} = \mathrm{ER}(p_k)$" into the model to obtain

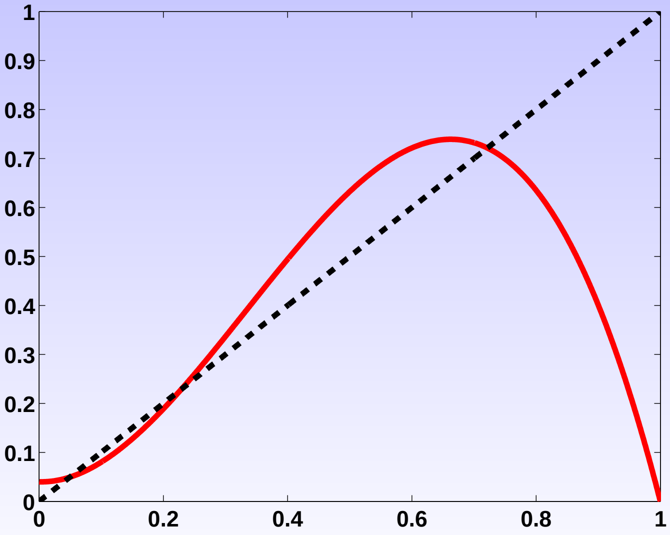
$$p_{k+1} = p_k(1-\omega) + (1-p_k)(\delta + \epsilon(n-2)p_k^2)$$

Generically: **three real roots**

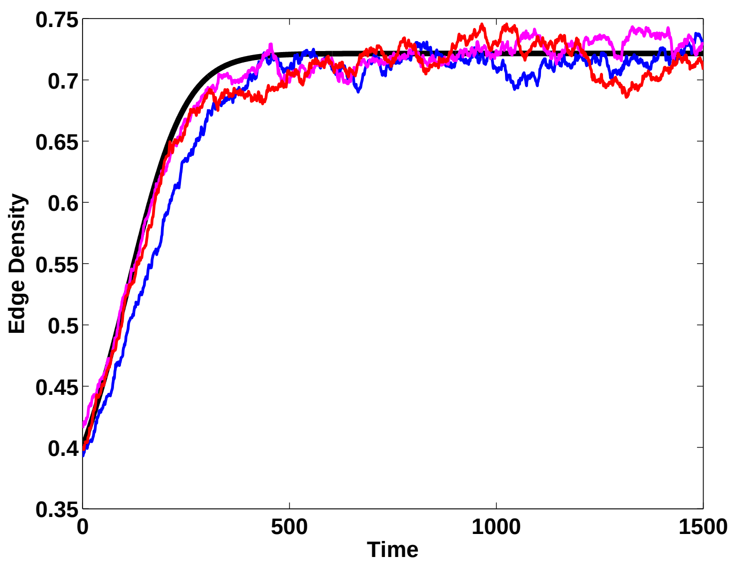
Two are stable, **one is unstable**

$n = 100$, $\omega = 0.01$, $\epsilon = 5 \times 10^{-4}$, $\delta = 4 \times 10^{-4}$
Stable fixed points 0.049 & 0.721 Unstable 0.229

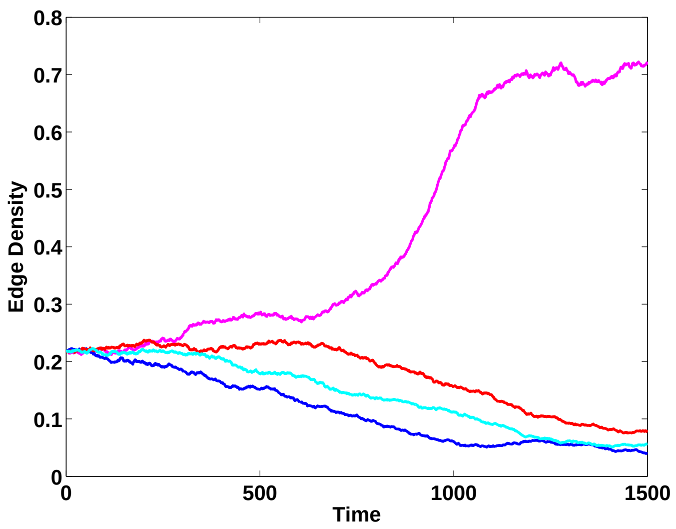
Fixed points 0.049, 0.721 and 0.229



Mean-field vs. simulation from $\mathrm{ER}(0.4)$



Four simulations from $\mathrm{ER}(0.23)$



Stable fixed points 0.049 & 0.721   Unstable 0.229

Calibration/Inference

Likelihood, $\mathcal{L}(A^{[k+1]}|A^{[k]})$, has the form

$$\prod_{\text{remain alive}} (1 - \text{death}) \prod_{\text{born}} \text{birth} \prod_{\text{remain dead}} (1 - \text{birth}) \prod_{\text{die}} \text{death}$$

Then the sequence $A^{[1]}, A^{[2]}, A^{[3]}, \ldots, A^{[K]}$ has likelihood

$$\mathcal{L}(A^{[1]}|A^{[0]})\,\mathcal{L}(A^{[2]}|A^{[1]})\,\mathcal{L}(A^{[3]}|A^{[2]}) \cdots \mathcal{L}(A^{[K-1]}|A^{[K]})$$

Constrained model with $\epsilon = 0$ is nested within the **unconstrained model**. We used a likelihood ratio test, and also computed the Akaike information criterion (AIC)

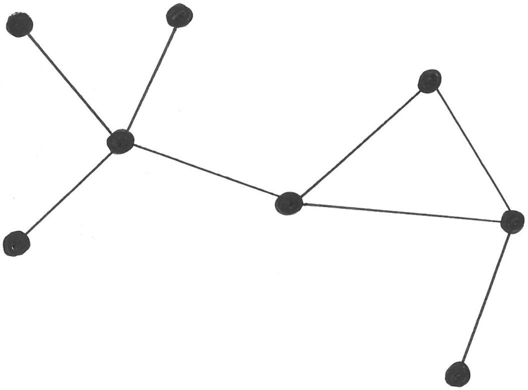
Tests on synthetic data show that we can correctly infer the triadic closure effect and recover a good estimate for ϵ

On **Wealink** data from Hu and Wang, Phys. Lett. A, 2009 with 26 Million time stamps, over 841 days and 0.25 Million nodes (no edge death), we found statistical support for triadic closure

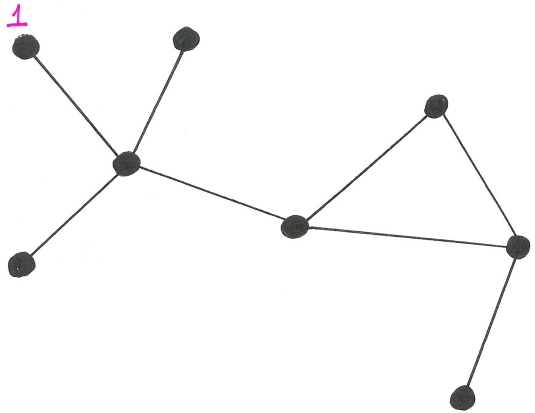
Part 2

The **Friendship Paradox**. . .

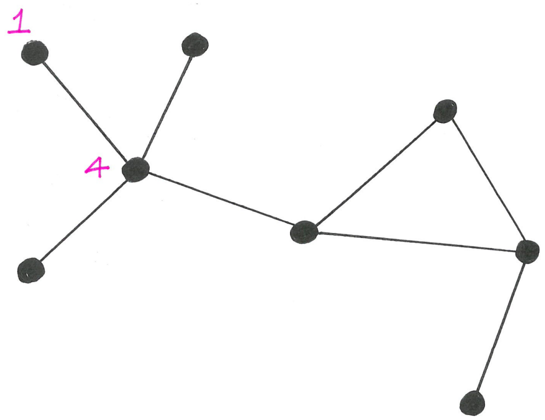
Example



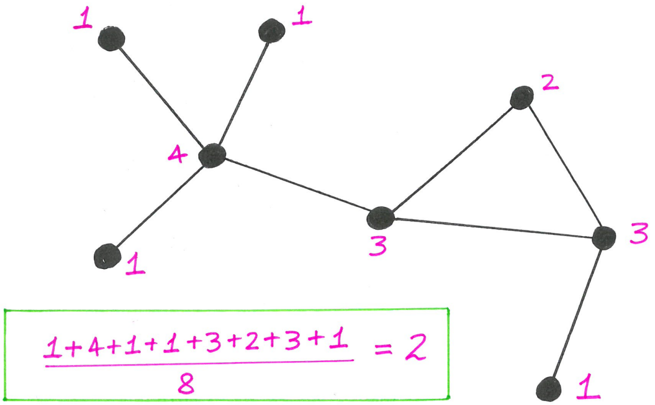
Let’s Count Average Num. Friends



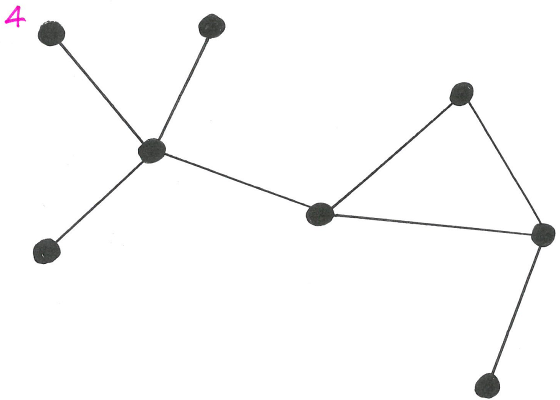
Let's Count Average Num. Friends



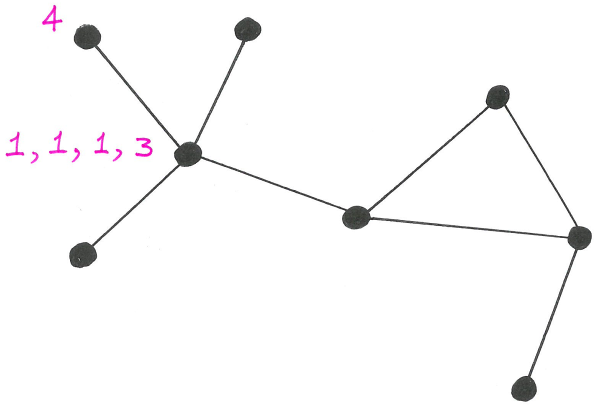
Let’s Count Average Num. Friends



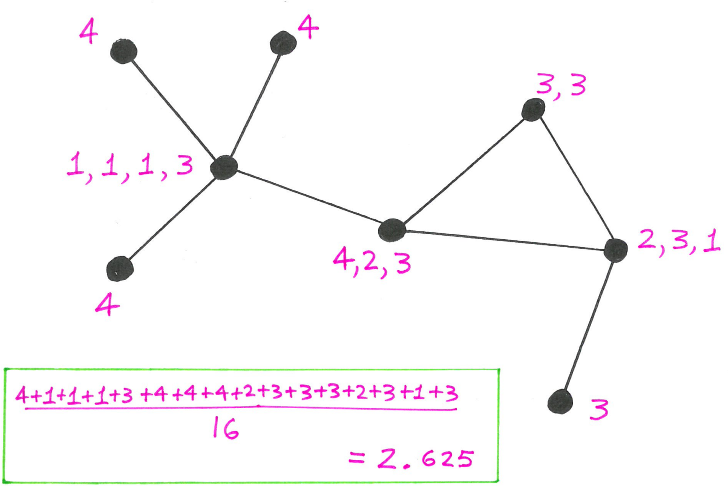
Average Num. of Friends of Friends



Average Num. of Friends of Friends



Average Num. of Friends of Friends



On average, our friends have more friends than we do

This is now called **The Friendship Paradox**

Why Your Friends Have More Friends Than You Do,
Scott L. Feld, The American Journal of Sociology, 1991

Quote: "**most individuals have friends who have more friends than average and so provide an unfair basis for comparison**"

We can blame the **Cauchy–Schwarz inequality**. . .

The Maths

Let $A \in \mathbb{R}^{n \times n}$ be the symmetric adjacency matrix
Let $\mathbf{d} = A\mathbf{1}$ be the degree vector

Average number of friends over the nodes is

$$\frac{1}{n}\sum_{i=1}^{n} d_i, \qquad \text{i.e.,} \qquad \frac{\|\mathbf{d}\|_1}{n}$$

Friend-of-friend average is

$$\frac{\sum_{i=1}^{n} d_i^2}{\sum_{i=1}^{n} d_i}, \qquad \text{i.e.,} \qquad \frac{\|\mathbf{d}\|_2^2}{\|\mathbf{d}\|_1}$$

The Maths

Let $A \in \mathbb{R}^{n \times n}$ be the symmetric adjacency matrix
Let $\mathbf{d} = A\mathbf{1}$ be the degree vector

Average number of friends over the nodes is

$$\frac{1}{n}\sum_{i=1}^{n} d_i, \qquad \text{i.e.,} \qquad \frac{\|\mathbf{d}\|_1}{n}$$

Friend-of-friend average is

$$\frac{\sum_{i=1}^{n} d_i^2}{\sum_{i=1}^{n} d_i}, \qquad \text{i.e.,} \qquad \frac{\|\mathbf{d}\|_2^2}{\|\mathbf{d}\|_1}$$

For $\mathbf{u}, \mathbf{v} \geq 0$, **Cauchy–Schwarz** $\Rightarrow \mathbf{u}^T\mathbf{v} \leq \|\mathbf{u}\|_2 \|\mathbf{v}\|_2$

The Maths

Let $A \in \mathbb{R}^{n \times n}$ be the symmetric adjacency matrix
Let $\mathbf{d} = A\mathbf{1}$ be the degree vector

Average number of friends over the nodes is

$$\frac{1}{n}\sum_{i=1}^{n} d_i, \qquad \text{i.e.,} \qquad \frac{\|\mathbf{d}\|_1}{n}$$

Friend-of-friend average is

$$\frac{\sum_{i=1}^{n} d_i^2}{\sum_{i=1}^{n} d_i}, \qquad \text{i.e.,} \qquad \frac{\|\mathbf{d}\|_2^2}{\|\mathbf{d}\|_1}$$

For $\mathbf{u}, \mathbf{v} \geq 0$, **Cauchy–Schwarz** $\Rightarrow \mathbf{u}^T\mathbf{v} \leq \|\mathbf{u}\|_2 \|\mathbf{v}\|_2$
Take $\mathbf{u} = \mathbf{1}$, so $\|\mathbf{v}\|_1 \leq \sqrt{n}\|\mathbf{v}\|_2$

The Maths

Let $A \in \mathbb{R}^{n \times n}$ be the symmetric adjacency matrix
Let $\mathbf{d} = A\mathbf{1}$ be the degree vector

Average number of friends over the nodes is

$$\frac{1}{n}\sum_{i=1}^{n} d_i, \qquad \text{i.e.,} \qquad \frac{\|\mathbf{d}\|_1}{n}$$

Friend-of-friend average is

$$\frac{\sum_{i=1}^{n} d_i^2}{\sum_{i=1}^{n} d_i}, \qquad \text{i.e.,} \qquad \frac{\|\mathbf{d}\|_2^2}{\|\mathbf{d}\|_1}$$

For $\mathbf{u}, \mathbf{v} \geq 0$, **Cauchy–Schwarz** $\Rightarrow \mathbf{u}^T\mathbf{v} \leq \|\mathbf{u}\|_2 \|\mathbf{v}\|_2$
Take $\mathbf{u} = \mathbf{1}$, so $\|\mathbf{v}\|_1 \leq \sqrt{n}\|\mathbf{v}\|_2 \Rightarrow \|\mathbf{v}\|_2^2 \geq \|\mathbf{v}\|_1^2/n$,

The Maths

Let $A \in \mathbb{R}^{n \times n}$ be the symmetric adjacency matrix
Let $\mathbf{d} = A\mathbf{1}$ be the degree vector

Average number of friends over the nodes is

$$\frac{1}{n}\sum_{i=1}^{n} d_i, \qquad \text{i.e.,} \qquad \frac{\|\mathbf{d}\|_1}{n}$$

Friend-of-friend average is

$$\frac{\sum_{i=1}^{n} d_i^2}{\sum_{i=1}^{n} d_i}, \qquad \text{i.e.,} \qquad \frac{\|\mathbf{d}\|_2^2}{\|\mathbf{d}\|_1}$$

For $\mathbf{u}, \mathbf{v} \geq 0$, **Cauchy–Schwarz** $\Rightarrow \mathbf{u}^T\mathbf{v} \leq \|\mathbf{u}\|_2 \|\mathbf{v}\|_2$
Take $\mathbf{u} = \mathbf{1}$, so $\|\mathbf{v}\|_1 \leq \sqrt{n}\|\mathbf{v}\|_2 \Rightarrow \|\mathbf{v}\|_2^2 \geq \|\mathbf{v}\|_1^2/n$, or

$$\frac{\|\mathbf{v}\|_2^2}{\|\mathbf{v}\|_1} \geq \frac{\|\mathbf{v}\|_1}{n}$$

Follow on Work

Paradox applies to **any mutual pairwise interactions**: minisymposium co-organisation, coauthorship, sexual partnership, . . .

Measured for many networks in social science, and **implications** extensively debated

A related idea has been used as a **sensing strategy**:

Social Network Sensors for Early Detection of Contagious Outbreaks,

Nicholas A. Christakis, James H. Fowler, PLoS ONE, 2010

Using Friends as Sensors to Detect Global-Scale Contagious Outbreaks,

Manuel Garcia-Herranz, Esteban Moro, Manuel Cebrian, Nicholas A. Christakis, James H. Fowler, PLoS ONE, 2014

Generalized Friendship Paradox

Generalized Friendship Paradox in Complex Networks: The case of scientific collaboration,

Young-Ho Eom, Hang-Hyun Jo, Scientific Reports, 2014

Do our friends have more of attribute **x** than us, on average?

E.g., for **scientific collaboration networks**, our coauthors seem to have more **citations** and **publications** than us, on average

They showed question boils down to $\mathrm{Cov}(\mathbf{x}, \mathbf{d}) \geq 0$? Equivalently

$$\frac{\mathbf{x}^T \mathbf{d}}{\|\mathbf{d}\|_1} \geq \frac{\|\mathbf{x}\|_1}{n}?$$

Always true when **x** is **eigenvector centrality**

Theorem: Eigenvector Centrality Paradox

For any connected graph the inequality

$$\frac{\mathbf{x}^T \mathbf{d}}{\|\mathbf{d}\|_1} \geq \frac{\|\mathbf{x}\|_1}{n}$$

holds when $\mathbf{x}$ is the **P-F vector** of A. We have equality if and only if the graph is regular.

Proof Let $A\mathbf{x} = \lambda \mathbf{x}$, with $\lambda = \rho(A)$. Then

$$\lambda = \| A \|_2 \geq \| A \frac{\mathbf{1}}{\sqrt{n}} \|_2 = \| \frac{\mathbf{d}}{\sqrt{n}} \|_2 \geq \frac{1}{n} \|\mathbf{d}\|_1.$$

Now

$$\frac{\mathbf{x}^T \mathbf{d}}{\|\mathbf{d}\|_1} = \frac{\mathbf{x}^T A \mathbf{1}}{\|\mathbf{d}\|_1} = \lambda \frac{\mathbf{x}^T \mathbf{1}}{\|\mathbf{d}\|_1} = \lambda \frac{\| \mathbf{x} \|_1}{\|\mathbf{d}\|_1} \geq \frac{\| \mathbf{x} \|_1}{n}.$$

$\Rightarrow$ **our friends are always at least as eigenvector central as us**, on average.

Triangle Paradox Inequality

Consider the case where x_i **counts the number of triangles that node i participates in**

Do we always have

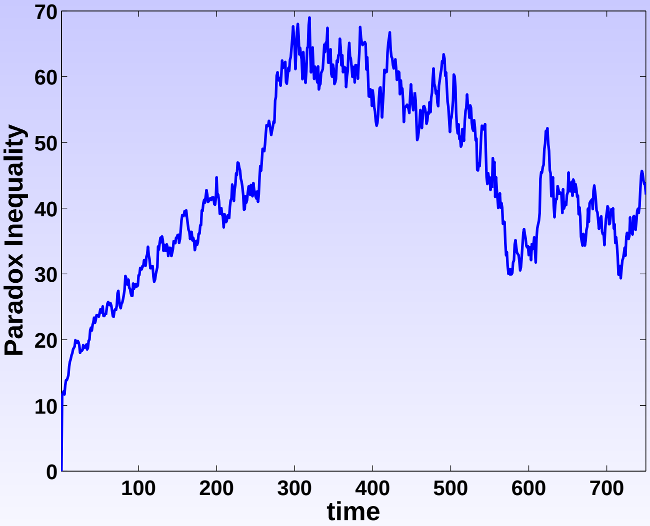
$$\frac{\mathbf{x}^T\mathbf{d}}{\|\mathbf{d}\|_1} - \frac{\|\mathbf{x}\|_1}{n} \geq 0 \quad \text{for } x_i = (A^3)_{ii}?$$

Not true in general

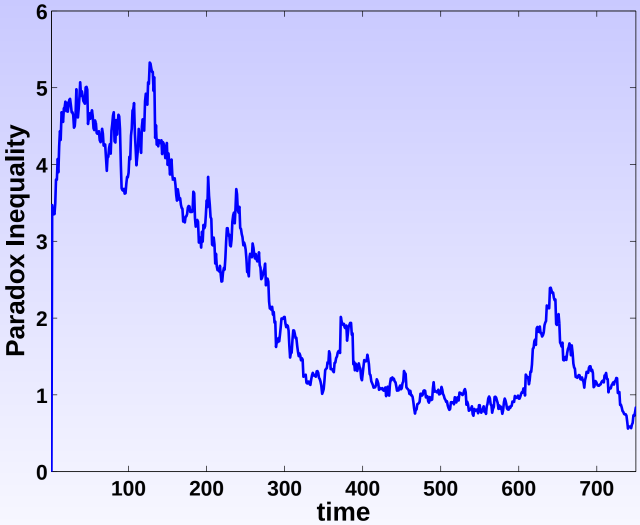
Related open question: **when does adding an edge make the LHS larger?**

See how the LHS evolves under the Markov chain triadic closure model. . .

Triangle Paradox Inequality from $\mathrm{ER}(0.3)$



Triangle Paradox Inequality from $\mathrm{ER}(0.15)$



Summary

- **Edge-independent dynamic networks** form a useful class of Markov chain models that can incorporate hypotheses from application areas
- **Triadic closure** model has cubic nonlinearity that leads to **bistable** behaviour
- Closing triangles over time can contribute to a **Triangle Paradox Inequality**
- Many opportunites for **further analysis**