Limiting Behavior in Large Networks

PJ Wolfe

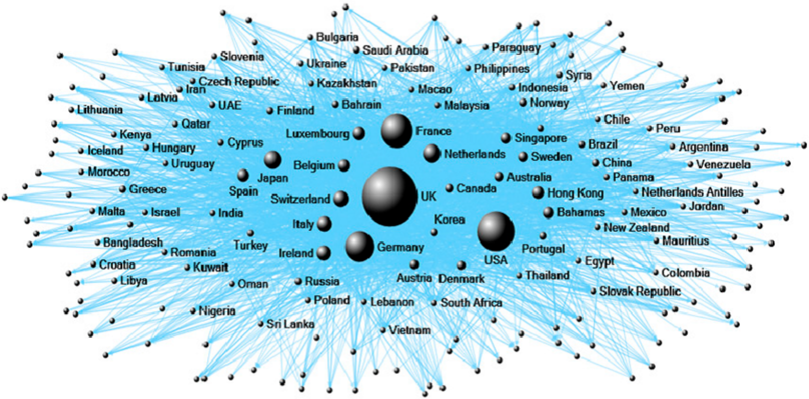
Chair of Statistics
Royal Society Research Fellow
University College London

IISc Workshop on High Dimensional Network Analytics
Bangalore, India, 19 December 2013

Fundamental question

How do we draw conclusions from and about networks?

Figure 1. The Global Banking Network

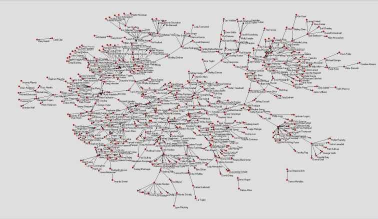


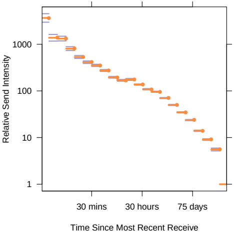
IMF Stability Assessment: Size + Interconnectedness $\Rightarrow$ "Importance"?

Statistical network modeling allows us to understand big data

Ingredients

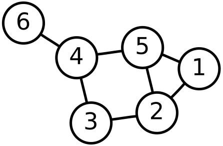
1. Mechanisms that generate data
2. Structure that facilitates analysis
3. Tools that can be understood

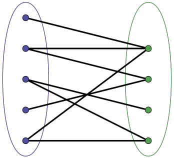




Modeling (big) networks

Objects + relationships ⇒ networks





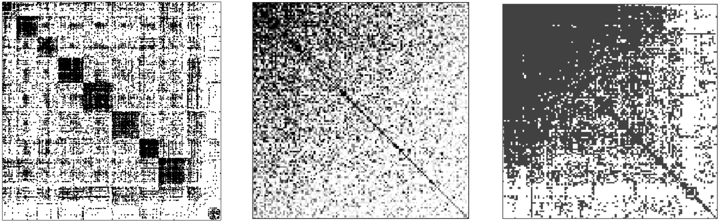
A network is two sets:

- Nodes ('nouns'): any discrete set of objects
- Links ('verbs'): a set of pairs of these objects

Examples:

- Social networks: people, friendships
- Complex systems: variables, correlations
- Images: pixel locations, similarities

Networks as (big) data



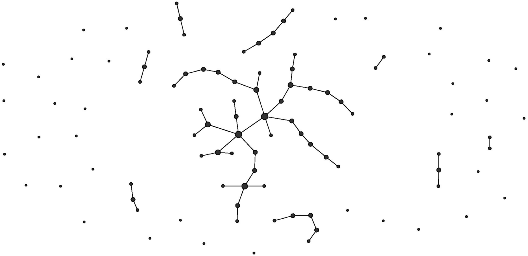
Adjacency matrices enable visualization of large networks:

$$(i,j)\text{th entry} = \begin{cases} 1 & \text{if node } i \text{ links to node } j, \\ 0 & \text{otherwise.} \end{cases}$$

Their structure reveals information:

- Links as (biased) coin tosses
- How to couple the tosses?

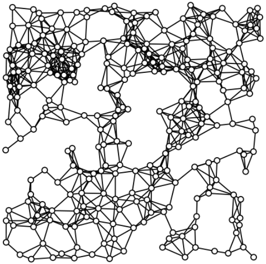
The simplest random network



Decoupled coin tosses

- Links as repeated tosses of the same coin, $\Pr(\text{Heads} = \frac{1}{100})$
- How many connections will form as the network grows?
- Emergence of triangles or other structure?

Clusters and connectedness



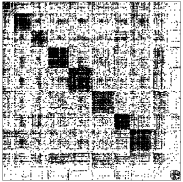
Strongly coupled coin tosses

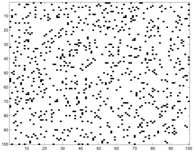
- Random scattering of nodes
- Nodes connect to others nearby
- As the network grows, will it eventually become connected?

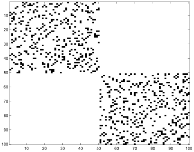
Limiting behavior in large networks

Cluster models

- Suppose n network nodes divide into k 'regular' groups (Szemerédi, Gowers, Tao)
- Thm: If k grows like $\sqrt{n}$, with edges added faster than n grows, then we can recover groups



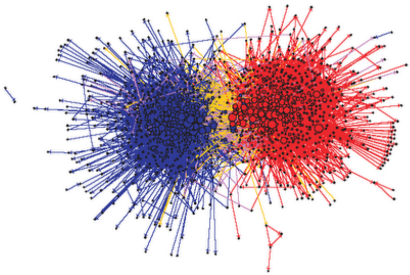




Graph limits

- Suppose a network is like a noisy 'image' f in the infinite limit (digital $\to$ analog)
- Thm: As $n \to \infty$, network yields 'oracle' information on f at a rate of at least $n^{-1/4}$
- Brute force algorithms needed to reveal this information... but many special cases possible

Making sense of clusters

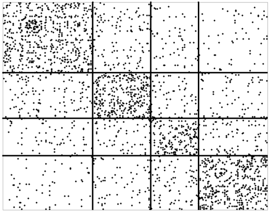


Political blog network (Adamic, 2005)

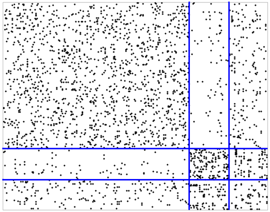
- What is the network equivalent of clustering?
- If the data are not generated by a cluster model, can we still approximate the generating mechanism (nonparametrics)?
- How to establish correct interpretation for nonparametrics?

Example: likelihood-based clustering

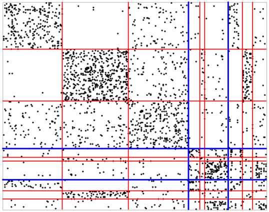
Survey data on high-school friendships (Add Health, 1994):



students grouped by year (black lines)



students grouped by race (blue lines)



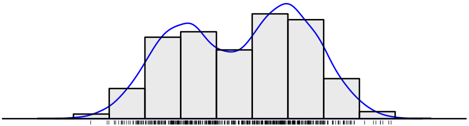
by race (blue lines) and then clustered (red)

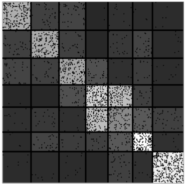
Implication: how do we interpret the notion of "clusters"?

Main result

- Assume the network to be generated nonparametrically, and then fit via likelihood-based clustering
- The resulting estimate approaches an optimal piecewise-constant approximation to the generative model

Implication: significantly broadens the interpretation of "clusters":

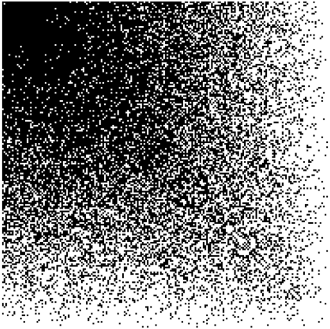




Generative model

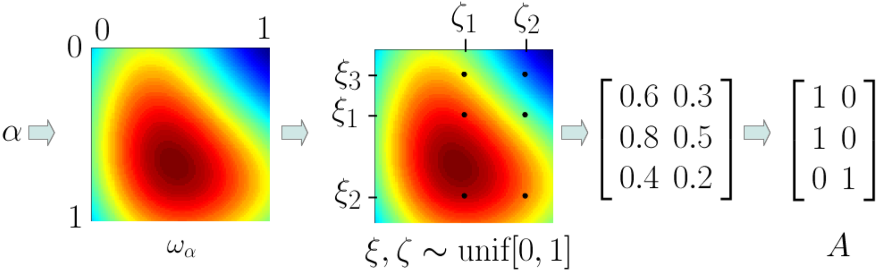
Example: Lovasz, *Very Large Graphs* (AMS, 2012):

Generative model



Adjacency matrix

Formal description

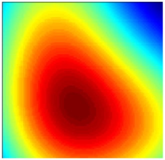


1. Specify $\underset{\sim}{\omega} : [0,1]^3 \to [0,1]$
2. Generate latent variable $\alpha \sim \text{Uniform}(0,1)$
3. Generate latent variables $\xi = (\xi_1, \ldots, \xi_m)$ and $\zeta = (\zeta_1, \ldots, \zeta_n) \sim \text{Uniform}(0,1)$
4. Let A_{ij} be Bernoulli with parameter $\omega_\alpha \equiv \underset{\sim}{\omega}(\alpha, \xi_i, \zeta_j)$. Connect nodes i and j if $A_{ij} = 1$.

Approximating the generative model

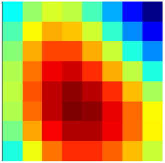
Qn: Can we fit a piecewise-constant approximation, ω_ϕ, to ω_α?

Given $\underset{A}{\begin{bmatrix} 1 & 0 & \dots \\ 1 & 0 & \\ \vdots & & \end{bmatrix}}$ generated by



ω_α

, estimate



ω_ϕ

Parameters $\phi \equiv (\mu, \nu, \theta)$ describe ω_ϕ:

- Vectors μ, ν : boundaries of the piecewise-constant regions
- Matrix θ: heights of the piecewise-constant regions

Fitting criteria

We can fit $\phi = (\mu, \nu, \theta)$ to an observed adjacency matrix A by various criteria:

$$\text{Likelihood: } L_A(\mu, \nu, \theta) = \max_{S,T} \sum_{i,j} \log \mathbb{P}(A_{ij} \mid \theta_{S(i)T(j)})$$

$$\text{Mean-squared error: } R_A(\mu, \nu, \theta) = \min_{S,T} \sum_{i,j} \left| A_{ij} - \theta_{S(i)T(j)} \right|^2$$

Mappings S and T assign nodes to K clusters, and are constrained to have assignment proportions matching μ and ν.

Risk measures

Fitting criteria correspond to risk functionals that measure agreement with an equivalence class induced by the unknown ω.

Let Π be the set of all measure-preserving bijective maps of $[0,1]$ to itself. Then we can minimize mean-squared error:

$$R_\omega(\phi) = \inf_{\pi_1,\pi_2\in\Pi} \iint_{(0,1)^2} \left|\omega\left(\pi_1(x),\pi_2(y)\right) - \omega_\phi(x,y)\right|^2 \, dx\, dy,$$

or maximize the corresponding likelihood as

$$\begin{aligned} L_\omega(\phi) = \sup_{\pi_1,\pi_2\in\Pi} \iint_{(0,1)^2} & \left[\omega\left(\pi_1(x),\pi_2(y)\right)\log\omega_\phi(x,y)\right. \\ & \left. + \left\{1-\omega\left(\pi_1(x),\pi_2(y)\right)\right\}\log\left\{1-\omega_\phi(x,y)\right\}\right] \, dx\, dy. \end{aligned}$$

Technical result

For the least squares co-blockmodel M-estimator

$$\hat{\phi} = \operatorname*{argmin}_{\phi \in \Phi} \left\{ \min_{S \in \mathcal{Q}_\mu^m, T \in \mathcal{Q}_\nu^n} \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} \left| \theta_{S(i)T(j)} - A_{ij} \right|^2 \right\}$$

relative to the L^2 risk

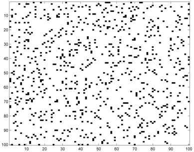
$$R_\omega(\phi) = \sup_{\pi_1, \pi_2 \in \Pi} \iint_{(0,1)^2} \left| \omega\left(\pi_1(x), \pi_2(y)\right) - \omega_\phi(x, y) \right|^2 \, dx \, dy,$$

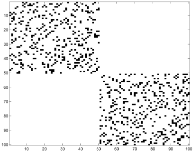
we have that

$$R_\omega(\hat{\phi}) - \inf_{\phi \in \Phi} R_\omega(\phi) = \mathcal{O}_P\left(n^{-1/4}\right).$$

An analogous result holds for $\sup_{\phi \in \Phi} L_\omega(\phi) - L_\omega(\hat{\phi})$.

Recap (Biometrika, 2012; Ann. Statist., 2013)



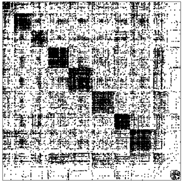


Graph limits

- Suppose a network is like a noisy 'image' f in the infinite limit (digital $\to$ analog)
- Thm: As $n \to \infty$, network yields 'oracle' information on f at a rate of at least $n^{-1/4}$
- Brute force algorithms needed to reveal this information... but many special cases possible

Cluster models

- Suppose n network nodes divide into k 'regular' groups (Szemerédi, Gowers, Tao)
- Thm: If k grows like $\sqrt{n}$, with edges added faster than n grows, then we can recover groups



Questions?

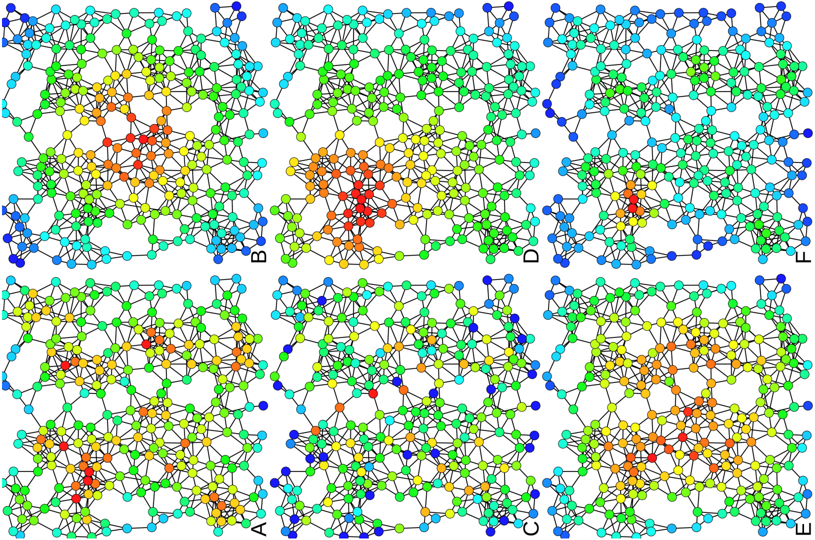


Image credits

Page 2: Figure 1 from *Integrating Stability Assessments Under the Financial Sector Assessment Program into Article IV Surveillance: Background Material*, Prepared by the Monetary and Capital Markets Department, August 27, 2010.

Page 3: www1.cs.columbia.edu/∼sara/cs6998/enron11-29-00m10.jpg; www.carlmaples.com/Enron_building_1893.html.

Page 6: Miller, B. A., Bliss, N. T., and Wolfe P. J., "Toward signal processing theory for graphs and non-Euclidean data," in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (2010).

Page 8: Newman, M. E. J. "Communities, modules and large-scale structure in networks." Nature Physics (2011): 25-31; Faculty of Physics and Astronomy, U. Würzberg, www.physik.uni-wuerzburg.de/typo3temp/pics/858ca46225.jpg; Copyright Channel Four Television Corporation

Page 9: http://en.wikipedia.org/wiki/File:Centrality.svg