Adventures in random graphs: Models, structures and algorithms

Armand M. Makowski

ECE & ISR/HyNet
University of Maryland at College Park
armand@isr.umd.edu

Complex networks

- Many examples
  - Biology (Genomics, protonomics)
  - Transportation (Communication networks, Internet, roads and railroads)
  - Information systems (World Wide Web)
  - Social networks (Facebook, LinkedIn, etc)
  - Sociology (Friendship networks, sexual contacts)
  - Bibliometrics (Co-authorship networks, references)
  - Ecology (food webs)
  - Energy (Electricity distribution, smart grids)
- Larger context of "Network Science"

Objectives

- Identify generic structures and properties of “networks
- Mathematical models and their analysis
- Understand how network structure and processes on networks interact

---

What is new?

Very large data sets now easily available!

- Dynamics **of** networks vs. dynamics **on** networks

Bibliography (I): Random graphs

- N. Alon and J.H. Spencer, *The Probabilistic Method* (Second Edition), Wiley-Science Series in Discrete Mathematics and Optimization, John Wiley & Sons, New York (NY) 2000.
- A. D. Barbour, L. Holst and S. Janson, *Poisson Approximation*, Oxford Studies in Probability **2**, Oxford University Press, Oxford (UK), 1992.
- B. Bollobás, *Random Graphs*, Second Edition, Cambridge Studies in Advanced Mathematics, Cambridge University Press, Cambridge (UK), 2001.
- R. Durrett, *Random Graph Dynamics*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge (UK), 2007.

- M. Draief and L. Massoulié, *Epidemics and Rumours in Complex Networks*, Cambridge University Press, Cambridge (UK), 2009.
- D. Dubhashi and A. Panconesi, *Concentration of Measure for the Analysis of Randomized algorithms*, Cambridge University Press, New York (NY), 2009.
- M. Franceschetti and R. Meester, *Random Networks for Communication: From Statistical Physics to Information Systems*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge (UK), 2007.
- S. Janson, T. Łuczak and A. Ruciński, *Random Graphs*, Wiley-Interscience Series in Discrete Mathematics and Optimization, John Wiley & Sons, 2000.
- R. Meester and R. Roy, *Continuum Percolation*, Cambridge University Press, Cambridge (UK), 1996.

- M.D. Penrose, *Random Geometric Graphs*, Oxford Studies in Probability **5**, Oxford University Press, New York (NY), 2003.

Bibliography (II): Survey papers

- R. Albert and A.-L. Barabási, "Statistical mechanics of complex networks," *Review of Modern Physics* **74** (2002), pp. 47-97.
- M.E.J. Newman, "The structure and function of complex networks," *SIAM Review* **45** (2003), pp. 167-256.

Bibliography (III): Complex networks

- A. Barrat, M. Barthelemy and A. Vespignani, *Dynamical Processes on Complex Networks*, Cambridge University Press, Cambridge (UK), 2008.
- R. Cohen and S. Havlin, *Complex Networks - Structure, Robustness and Function*, Cambridge University Press, Cambridge (UK), 2010.
- D. Easley and J. Kleinberg, *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*, Cambridge University Press, Cambridge (UK) (2010).
- M.O. Jackson, *Social and Economic Networks*, Princeton University Press, Princeton (NJ), 2008.

- M.E.J. Newman, A.-L. Barabási and D.J. Watts (Editors), *The Structure and Dynamics of Networks*, Princeton University Press, Princeton (NJ), 2006.

LECTURE 1

Basics of graph theory

What are graphs?

With V a finite set, a graph G is an ordered pair (V, E) where elements in V are called **vertices/nodes** and E is the set of **edges/links**:

$$E \subseteq V \times V$$

$$\mathcal{E}(G) = E$$

Nodes i and j are said to be **adjacent**, written $i \sim j$, if

$$e = (i, j) \in E, \quad i, j \in V$$

Multiple representations for $G = (V, E)$

Set-theoretic – **Edge variables** $\{\xi_{ij},\ i, j \in V\}$ with

$$\xi_{ij} = \begin{cases} 1 & \text{if } (i,j) \in E \\ \\ 0 & \text{if } (i,j) \notin E \end{cases}$$

Algebraic – **Adjacency matrix** $A = (A_{ij})$ with

$$A_{ij} = \begin{cases} 1 & \text{if } (i,j) \in E \\ \\ 0 & \text{if } (i,j) \notin E \end{cases}$$

Some terminology

Simple graphs vs. multigraphs

Directed vs. **undirected**

$$(i,j) \in E \quad \text{if and only if} \quad (j,i) \in E$$

No self loops

$$(i,i) \notin E, \quad i \in V$$

Here: Simple, undirected graphs with no self loops!

Types of graphs

- The empty graph
- Complete graphs
- Trees/forests
- A **subgraph** $H = (W, F)$ of $G = (V, E)$ is a graph with vertex set W such that

$$W \subseteq V \quad \text{and} \quad F = E \cap (W \times W)$$

- Cliques (Complete subgraphs)

Labeled vs. unlabeled graphs

A graph **automorphism** of $G = (V, E)$ is any one-to-one mapping $\sigma : V \to V$ that preserves the graph structure, namely

$$(\sigma(i), \sigma(j)) \in E \quad \text{if and only if} \quad (i, j) \in E$$

Group $\text{Aut}(G)$ of graph automorphisms of G

Of interest

- Connectivity and k-connectivity (with $k \geq 1$)
- Number and size of components
- Isolated nodes
- Degree of a node: degree distribution/average degree, maximal/minimal degree
- Distance between nodes (in terms of number of hops): Shortest path, diameter, eccentricity, radius
- Small graph containment (e.g., triangles, trees, cliques, etc.)
- Clustering
- Centrality: Degree, closeness, in-betweenness

For i, j in V,

$$\ell_{ij} = \begin{array}{c}\text{Shortest path length between} \\ \text{nodes } i \text{ and } j \text{ in the graph } G = (V, E)\end{array}$$

Convention: $\ell_{ij} = \infty$ if nodes i and j belong to different components and $\ell_{ii} = 0$.

Average distance

$$\ell_{\text{Avg}} = \frac{1}{|V|(|V| - 1)} \sum_{i \in V} \sum_{j \in V} \ell_{ij}$$

Diameter

$$d(G) = \max\left(\ell_{ij},\ i, j \in V\right)$$

Eccentricity

$$\mathrm{Ec}(i) = \max\left(\ell_{ij},\ j \in V\right), \quad i \in V$$

Radius

$$\mathrm{rad}(G) = \min\left(\mathrm{Ec}(i),\ i \in V\right)$$

$$\ell_{\mathrm{Avg}} \leq d(G)$$

and

$$\mathrm{rad}(G) \leq d(G) \leq 2\ \mathrm{rad}(G)$$

Centrality

Q: How central is a node?

Closeness centrality

$$g(i) = \frac{1}{\sum_{j \in V} \ell_{ij}}, \quad i \in V$$

Betweenness centrality

$$b(i) = \sum_{k \neq i,\ k \neq j,} \frac{\sigma_{kj}(i)}{\sigma_{kj}}, \quad i \in V$$

Clustering

Clustering coefficient of node i

$$C(i) = \frac{\sum_{j \neq i,\ k \neq i,\ j \neq k} \xi_{ij}\xi_{ik}\xi_{kj}}{\sum_{j \neq i,\ k \neq i,\ j \neq k} \xi_{ij}\xi_{ik}}$$

Average clustering coefficient

$$C_{\text{Avg}} = \frac{1}{n} \sum_{i \in V} C(i)$$

$$C = 3 \cdot \frac{\text{Number of fully connected triples}}{\text{Number of triples}}$$

Random graphs

Random graphs?

$\mathcal{G}(V) \equiv$ Collection of all (simple free of self-loops undirected) graphs with vertex set V.

Definition – Given a probability triple $(\Omega, \mathcal{F}, \mathbb{P})$, a **random graph** is simply a graph-valued rv $\mathbb{G} : \Omega \to \mathcal{G}(V)$.

Modeling – We need only specify the **pmf**

$$\{\mathbb{P}[\mathbb{G} = G], \quad G \in \mathcal{G}(V)\}.$$

Many, many ways to do that!

Equivalent representations for $\mathbb{G}$

Set-theoretic – **Link assignment rvs** $\{\xi_{ij},\ i,j \in V\}$ with

$$\xi_{ij} = \begin{cases} 1 & \text{if } (i,j) \in \mathcal{E}(\mathbb{G}) \\ \\ 0 & \text{if } (i,j) \notin \mathcal{E}(\mathbb{G}) \end{cases}$$

Algebraic – Random **adjacency matrix** $A = (A_{ij})$ with

$$A_{ij} = \begin{cases} 1 & \text{if } (i,j) \in \mathcal{E}(\mathbb{G}) \\ \\ 0 & \text{if } (i,j) \notin \mathcal{E}(\mathbb{G}) \end{cases}$$

Why random graphs?

Useful models in many applications to capture binary relationships between participating entities

---

Because

$$\begin{aligned} |\mathcal{G}(V)| &= 2^{\frac{|V|(|V|-1)}{2}} \\ &\simeq 2^{\frac{|V|^2}{2}} \quad \text{A very large number!} \end{aligned}$$

there is a need to identify/discover **typicality**!

Scaling laws – Zero-one laws as $|V|$ becomes large, e.g.,

$$V \equiv V_n = \{1, \ldots, n\} \quad (n \to \infty)$$

Ménagerie of random graphs

- Erdős-Renyi graphs $\mathbb{G}(n; m)$
- Erdős-Renyi graphs $\mathbb{G}(n; p)$
- Generalized Erdős-Renyi graphs
- Geometric random models/disk models
- Intrinsic fitness and threshold random models
- Random intersection graphs
- Growth models: Preferential attachment, copying
- Small worlds
- Exponential random graphs
- Etc

Erdős-Renyi graphs $\mathbb{G}(n;m)$

With

$$1 \leq m \leq \binom{n}{2} = \frac{n(n-1)}{2},$$

the pmf on $\mathcal{G}(V_n)$ is specified by

$$\mathbb{P}\left[\mathbb{G}(n;m) = G\right] = \begin{cases} u(n;m)^{-1} & \text{if } |\mathcal{E}(G)| = 2m \\ \\ 0 & \text{if } |\mathcal{E}(G)| \neq 2m \end{cases}$$

where

$$u(n;m) = \binom{\frac{n(n-1)}{2}}{m}$$

Uniform selection over the collection of all graphs on the vertex set $\{1, \ldots, n\}$ with exactly m edges

Erdős-Renyi graphs $\mathbb{G}(n; p)$

With

$$0 \le p \le 1,$$

the link assignment rvs $\{\chi_{ij}(p),\ 1 \le i < j \le n\}$ are **i.i.d.** $\{0, 1\}$-valued rvs with

$$\mathbb{P}\left[\chi_{ij}(p) = 1\right] = 1 - \mathbb{P}\left[\chi_{ij}(p) = 0\right] = p, \qquad 1 \le i < j \le n$$

For every G in $\mathcal{G}(V)$,

$$\mathbb{P}\left[\mathbb{G}(n; p) = G\right] = p^{\frac{|\mathcal{E}G|}{2}} \cdot (1 - p)^{\frac{n(n-1)}{2} - \frac{|\mathcal{E}G|}{2}}$$

Related to, but easier to implement than $\mathbb{G}(n;m)$

Similar behavior/results under the **matching condition**

$$|\mathcal{E}(\mathbb{G}(n;m))| = \mathbb{E}\left[|\mathcal{E}(\mathbb{G}(n;p))|\right],$$

namely

$$m = \frac{n(n-1)}{2}p$$

Generalized Erdős-Renyi graphs

With

$$0 \leq p_{ij} \leq 1, \quad 1 \leq i < j \leq n$$

the link assignment rvs $\{\chi_{ij}(p),\ 1 \leq i < j \leq n\}$ are **mutually independent** $\{0,1\}$-valued rvs with

$$\mathbb{P}\left[\chi_{ij}(p_{ij}) = 1\right] = 1 - \mathbb{P}\left[\chi_{ij}(p_{ij}) = 0\right] = p_{ij}, \quad 1 \leq i < j \leq n$$

An important case: With positive weights $w_1, \ldots, w_n$,

$$p_{ij} = \frac{w_i w_j}{W} \quad \text{with} \quad W = w_1 + \ldots + w_n$$

Geometric random graphs $(d \geq 1)$

With **random** locations in $\mathbb{R}^d$ at

$$\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n,$$

the link assignment rvs $\{\chi_{ij}(\rho),\ 1 \leq i < j \leq n\}$ are given by

$$\chi_{ij}(\rho) = \mathbf{1}\left[\ \|\boldsymbol{X}_i - \boldsymbol{X}_j\| \leq \rho\ \right], \qquad 1 \leq i < j \leq n$$

where $\rho > 0$.

Usually, the rvs $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ are taken to be **i.i.d.** rvs **uniformly** distributed over some compact subset $\Gamma \subseteq \mathbb{R}^d$

Even then, not so obvious to write

$$\mathbb{P}\left[\mathbb{G}(n;\rho) = G\right], \quad G \in \mathcal{G}(V).$$

since the rvs $\{\chi_{ij}(\rho),\ 1 \leq i < j \leq n\}$ are no more i.i.d. rvs

---

For $d = 2$, long history for modeling wireless networks (known as the **disk model**) where ρ interpretated as transmission range

---

Threshold random graphs

Given $\mathbb{R}_+$-valued **i.i.d.** rvs $W_1, \ldots, W_n$ with absolutely continuous probability distribution function F,

$$i \sim j \quad \text{if and only if} \quad W_i + W_j > \theta$$

for some $\theta > 0$

Generalizations:

$$i \sim j \quad \text{if and only if} \quad R(W_i, W_j) > \theta$$

for some symmetric mapping $R : \mathbb{R}_+^2 \to \mathbb{R}_+$.

Random intersection graphs

Given a **finite** set $\mathcal{W} \equiv \{1, \ldots, W\}$ of features, with **random** subsets $K_1, \ldots, K_n$ of $\mathcal{W}$,

$$i \sim j \quad \text{if and only if} \quad K_i \cap K_j \neq \emptyset$$

Co-authorship networks, random key distribution schemes, classification/clustering

Growth models

$$\{\mathbb{G}_t,\ t = 0, 1, \ldots\}$$

with rules

$$V_{t+1} \leftarrow V_t$$

and

$$\mathbb{G}_{t+1} \leftarrow (\mathbb{G}_t, V_{t+1})$$

- Preferential attachment
- Copying

Scale-free networks

Small worlds

- Between randomness and order
- Shortcuts
- Short paths but high clustering

Milgram's experiment and six degrees of separation

Exponential random graphs

- Models favored by sociologists and statisticians
- Graph analog of **exponential** families often used in statistical modeling
- Related to Markov random fields

With I parameters

$$\boldsymbol{\theta} = (\theta_1, \ldots, \theta_I)$$

and a set of I observables (statistics)

$$u_i : \mathcal{G}(V) \to \mathbb{R}_+,$$

we postulate

$$\mathbb{P}\left[\mathbb{G} = G\right] = \frac{e^{\sum_{i=1}^{I} \theta_i u_i(G)}}{Z(\boldsymbol{\theta})}, \quad G \in \mathcal{G}(V)$$

with normalization constant

$$Z(\boldsymbol{\theta}) = \sum_{G \in \mathcal{G}(V)} e^{\sum_{i=1}^{I} \theta_i u_i(G)}$$

In sum

- Many different ways to specify the pmf on $\mathcal{G}(V)$
  - Local description vs. global representation
  - Static vs. dynamic
  - Application-dependent mechanisms