Random Graphs and Wireless Communication Networks

Part 1. An Introduction to Random Graph Theory and Network Science

September 5, 2016

Shahriar Etemadi Tajbakhsh

with Justin Coon, Carl Dettmann, Marco Di Renzo and Orestis Georgiou

Outline

- History
- Models of Random Graphs
- Random Geometric Graphs
- The Clustering Coefficient
- Fundamental Theorem
- Connectedness
- Small World
  - Watts and Strogatz Model
  - Preferential Attachment
- Centrality
- PageRank Algorithm
- Temporal Networks
- Complex Networks
- Other Methods in Complex Networks: Epidemic Diseases Networks

History

Euler in 1736 proved that it is not possible to devise a walk such that each bridge is crossed once and only once. Graph theory started!

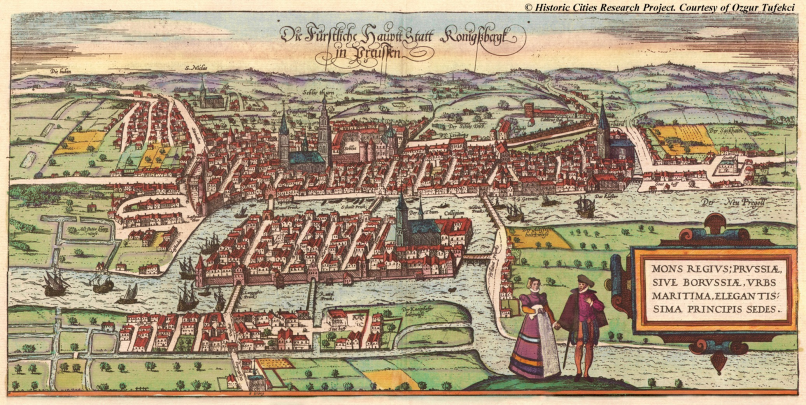


Figure: Seven Bridges of Königsberg

History (Cont'd)

Paul Erdös and Alfred Rényi, two Hungarian mathematicians introduced random graphs in 1959 (and also independently by Gilbert).

Figure: Budapest: Hometown of Erdös-Rényi

Models of Random Graphs

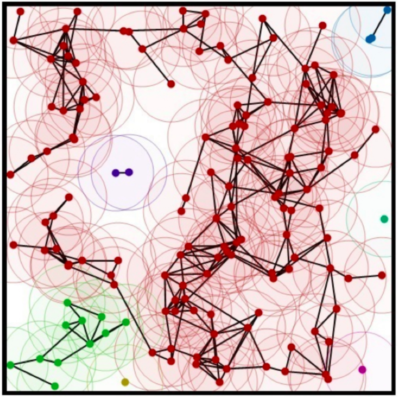
- *Model 1*: $\mathcal{G}(n, M)$: the model consists of all graphs with vertex set $V = \{1, 2, \ldots, n\}$ having M edges in which the graphs have the same probability.
  - Probability of each graph: $\binom{N}{M}^{-1}$ where $N = \binom{n}{2}$
- *Model 2*: $\mathcal{G}\{n, P(edge) = p\}, 0 < p < 1$, the model consists of all graphs with vertex set $V = \{1, 2, \ldots, n\}$ in which the edges are chosen independently and with probability p.
  - If G_0 is a graph with vertex set V and it has m edges then $P(\{G_0\}) = P(G = G_0) = p^m q^{N-m}$

Some generalized models:

- $\mathcal{G}\{n, p_{ij}\}$ for vertex set V and $1 \leqslant i < j \leqslant n$, in which edges are chosen independently where the probability of ij being an edge is p_{ij}.
  - *Special Case*: $\mathcal{G}(H; p)$ where H is a fixed graph and $0 < p < 1$. If $V(H) = V$: $p_{ij} = \begin{cases} p & \text{if } ij \in E(H), \\ 0 & \text{if } \textit{otherwise} \end{cases}$.

Random Geometric Graphs (RGG)

- A *random geometric graph* (RGG) $\mathcal{G}\{n, r\}$ consists of set of points randomly distributed in a d-dimensional space as its vertex set, where the probability of an edge existing between two vertices i and j depends on the Euclidean distance of them. In its simplest form, an edge exists between i and j if $||X_i - X_j|| \leqslant R$.

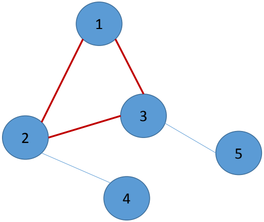


  - The main difference between RGG and Erdös-Rényi graph is that in RGG the existence of edges are not independent. If X_i is close to X_j and X_j close to X_k, then X_i is close to X_k.

The Clustering Coefficient

- Erdös-Rényi model is not the most precise model for many real world networks. Real world networks are complex. Nodes have a tendency to form a cluster, e.g. we have a circle friends where most people know each other in this circle.
- *Clustering coefficient* measures the degree to which the nodes in a network (graph) tend to cluster.

$$C = \frac{3 \times \text{number of triangles}}{\text{number of connected triplets of vertices}} = \frac{\text{number of closed triplets}}{\text{number of connected triplets}}$$



$$C = \frac{3 \times 1}{6} = \frac{1}{2}$$

Fundamental Theorem

How random graphs are studied?

Theorem

Let Q be any property and suppose $pqN \to \infty$. Then almost every graph in $\mathcal{G}(n, p)$ has Q.

Example

Theorem

If the size of maximum clique in a random graph $\mathcal{G}\{n, p\}$ is denoted by X_n and $n \to \infty$, then
$\frac{X_n}{\log n} \to \frac{2}{log(\frac{1}{p})}$

Connectedness

- *Giant Component*: A giant component is a connected component of a given random graph that contains a constant fraction of the entire graph's vertices.
- In a random graph $\mathcal{G}\{n, p\}$ if $p \leqslant \frac{1-\epsilon}{n}$ for any constant $\epsilon > 0$ then with high probability all the connected components have a size $\mathcal{O}(\log n)$.
- If $p \geqslant \frac{1+\epsilon}{n}$, w.h.p there is a giant component with all other components having size $\mathcal{O}(\log n)$.
- For $p = \frac{1}{n}$ the largest component is w.h.p proportional to $n^{\frac{2}{3}}$.

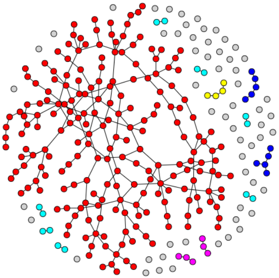


Figure: Components of a graph

Finding such critical points where a phase transition occurs in a random graph is the subject of *Percolation Theory*.

Small World Model

- A *small world* is a graph in which not many nodes are neighbors but most of them can reach each other with a small number of hops.
- Example: Six degrees of separation in human social networks. With how many steps are you related to Obama?
- Formal definition: A small world is a graph where the average distance between two randomly chosen nodes, denoted by L is proportional to the logarithm of the size of the network.
  $L \propto \log n$

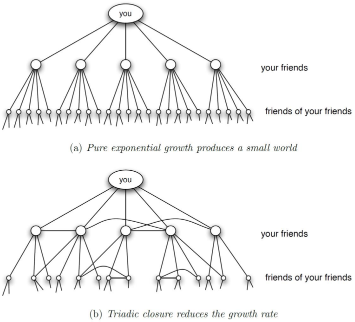


Figure: Formation of a small world (from book *Networks, Crowds, and Markets*, D. Easly and J. Kleinberg, 2010)

Watts and Strogatz Model

- Erdös-Rényi graphs do not typically generate small worlds, as the links between nodes are generated independently and this does not lead to *local clustering* and *triadic closure*. Moreover, in ER graphs the degree distribution tends to a Poisson distribution while in many real world networks the degree distribution is of power law.
- Algorithm
  1. Create a regular ring lattice with N nodes where each node is connected to $\frac{K}{2}$ nodes on each side.
  2. For every node n_i take every edge (n_i, n_j) and rewire it with probability β.

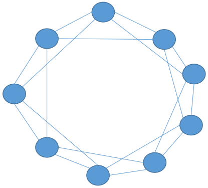


Figure: Regular Ring Lattice

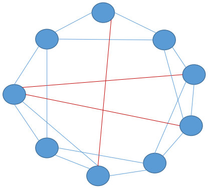


Figure: Small World

Preferential Attachment

- Rich gets richer!
- An initial set of $n_0 \geqslant 2$ nodes is considered, where the degree of each node is at least 1. A new node, with probaility p_i connects itself to node i.
  $p_i = \frac{d_i}{\sum_j d_j}$
  where d_j is the degree of node j.
- Nodes already with higher degree attract even more and more new nodes and nodes with smaller degree have less chance to attract new nodes.
- The degree distribution under preferential attachment model is shown to be free scale and follow a power law of the following form.

$$p(k) \sim k^{-3}$$

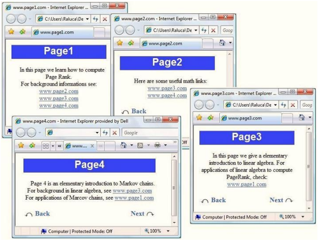
Centrality in Networks

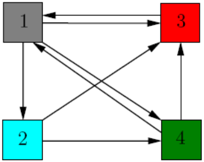
Nodes (and edges) in networks are of different importance levels!
Various measures for nodes' centrality exist in the literature.

- *Degree Centrality*: Nodes with higher degree are more susceptible (e.g. to viruses or receiving some information, etc.)
  $C_D(v) = deg(v)$
- *Closeness Centrality*: Measures the farness of a node from other nodes.
  $C_C(v) = \frac{1}{\sum_{u \in V} d(u,v)}$
- *Betweenness Centrality*: Measures how many shortest paths between two other nodes passes through this node (how critical a node is as a bridge!)
  $C_B(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$
  where $\sigma_{st}(v)$ is the number of shortest paths passing through v and σ_{st} is the total number of shortest paths between s and t.

PageRank Algorithm

Another interesting centrality measure is used and by Google to sort the search results based on their importance. This algorithm called PageRank algorithm was invented by Larry Page and Sergey Brin (The co-founders of Google) in 1998. A web page is considered to be *important* if many other pages or a few but important pages refer to it.





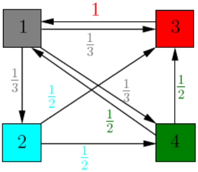


Figure: PageRank Algorithm From Cornell University Math Explorer Club Website

PageRank Algorithm (Cont'd)

Each page divides its importance equally among the pages it links and transfers it to them. This relationship is represented by a transition matrix A. Initially we assume that all the pages are of equal importance. Therefore, if we have k pages, each page has $\frac{1}{k}$ of importance. In our example the initial importance vector is $v = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$.

$$A = \begin{bmatrix} 0 & 0 & 1 & \frac{1}{2} \\ \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{2} & 0 & 0 \end{bmatrix}.$$

$$\mathbf{v} = \begin{pmatrix} 0.25 \\ 0.25 \\ 0.25 \\ 0.25 \end{pmatrix}, \quad \mathbf{Av} = \begin{pmatrix} 0.37 \\ 0.08 \\ 0.33 \\ 0.20 \end{pmatrix}, \quad \mathbf{A^2 v} = \mathbf{A\,(Av)} = \mathbf{A} \begin{pmatrix} 0.37 \\ 0.08 \\ 0.33 \\ 0.20 \end{pmatrix} = \begin{pmatrix} 0.43 \\ 0.12 \\ 0.27 \\ 0.16 \end{pmatrix}$$

$$\mathbf{A^3 v} = \begin{pmatrix} 0.35 \\ 0.14 \\ 0.29 \\ 0.20 \end{pmatrix}, \quad \mathbf{A^4 v} = \begin{pmatrix} 0.39 \\ 0.11 \\ 0.29 \\ 0.19 \end{pmatrix}, \quad \mathbf{A^5 v} = \begin{pmatrix} 0.39 \\ 0.13 \\ 0.28 \\ 0.19 \end{pmatrix}$$

$$\mathbf{A^6 v} = \begin{pmatrix} 0.38 \\ 0.13 \\ 0.29 \\ 0.19 \end{pmatrix}, \quad \mathbf{A^7 v} = \begin{pmatrix} 0.38 \\ 0.12 \\ 0.29 \\ 0.19 \end{pmatrix}, \quad \mathbf{A^8 v} = \begin{pmatrix} 0.38 \\ 0.12 \\ 0.29 \\ 0.19 \end{pmatrix}$$

Temporal Networks

- A temporal network consists of a set of vertices V where the presence or absence of any edges is a function of time, e.g., an edge might exist between two vertices i and j and may disappear at another time. An edge in a temporal network is identified by its two ends and its time stamp represented by the triplet (i, j, t).
- A path in a temporal network is different from a path in a static graph. A temporal path is a sequence of temporal edges $(i_1, i_2, t_1), (i_2, i_3, t_2), \ldots, (i_{n-1}, i_n, t_{n-1})$ where $t_1 < t_2 < \cdots < t_{n-1}$. Therefore the sequence of edges should also satisfy the *causality*. t_{n-1} is called the *duration* of this path.
- Many other aspects of temporal networks are different from static graphs in the same way as the temporal paths, for instance centrality, connectedness, etc. should be revised accordingly.

Accessibility (Reachability) Graph in Temporal Networks

There exist an edge between two vertices i and j in the accessibility graph at time slot T if there exists a temporal path between i and j within the period $1, \ldots, T$. If we denote the adjacency matrix of the temporal network at each time slot t by $\mathbf{A}_t$, then $C_T = \prod_{t=1}^{T}(\mathbf{1} + \mathbf{A}_t)$ gives the number of temporal paths between any two vertices. By changing any non-zero element in C_T to 1 the adjacency matrix of accessibility graph is obtained.

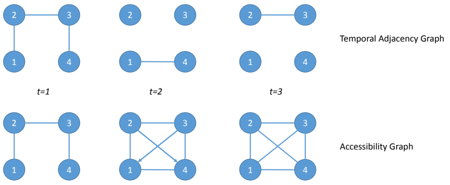


Figure: Formation of Accessibility Graph

The accessibility graph is directed in general. This is an immediate consequence of causality.

Complex Networks

- Social Human and Computer Networks
- Contact Based Networks: Epidemic Diseases
- Web Networks, Citation Networks, etc.
- Wireless Delay Tolerant Networks
- Biological Networks

Figure: Complex Networks

Other Methods in Complex Networks: Epidemic Diseases Networks

SIR model

S: number susceptible, I: number infected, R: number recovered

$\frac{dS}{dt} = -\beta IS$

$\frac{dI}{dt} = \beta IS - \gamma I$

$\frac{dR}{dt} = \gamma I$

$\frac{dS}{dt} + \frac{dI}{dt} + \frac{dR}{dt} = 0 \Longrightarrow S(t) + I(t) + R(t) = constant = N$ (size of population)