Networks - Week 2 - Exploring Properties of Networks

Antonio León Villares

October 2023

Contents

Based on the notes by Renaud Lambiotte and Pete Grindrod, Chapters 2 & 3

1 Properties of Networks

1.1 Definition: Network

> *A **network** G consists of a set of **vertices** V connected by **links** E:*
>
> $$G = (V, E)$$
>
> *An **edge** $e \in E$ is defined by a pair of vertices:*
>
> $$e = (v_i, v_j)$$

- **What is an undirected network?**
  - a network whereby if $(v_i, v_j) \in E$ then $(v_j, v_i) \in E$ (these typically just count as a single edge)
  - links don't have any "direction"
- **What is a directed network?**
  - a network in which links have directionality
  - if $(v_i, v_j), (v_j, v_i) \in E$ then v_i, v_j are said to be **reciprocally connected**
- **What is a weighted network?**
  - a network where edges have a **weight function** assigned
  - this denotes some property of the network

1.2 Definition: Adjacency Matrix

> *Let $G = (V, E)$ be a **network**. The **adjacency matrix** A of G is a $|V| \times |V|$ matrix given by:*
>
> $$A_{ij} = \begin{cases} 1, & (v_i, v_j) \in E \\ 0, & \text{otherwise} \end{cases}$$
>
> ---
>
> *Generally, G shouldn't contain **self connections** or **double edges**, so A is **binary** and has 0s along its **diagonal**.*

1.3 Definition: Walk

> *A **walk** through a network G is an **ordered sequence** of **links** such that the ending **vertex** of the ith edge is the starting **vertex** of the $i+1$th edge.*

- **What is the length of a walk?**
  - the number of separate sequential edges

1.4 Definition: Path

> *A **path** is a **walk** where each vertex in the network is visited **only once**.*
>
> ---
>
> *If the **walk** ends in the **vertex** where it started, the **path** is called a **cycle**.*

- **In what scenarios are walks used?**
  - typically to emulate dynamical processes (i.e random walks)
- **In what scenarios are paths used?**
  - typically to consider the shortest travelling route between vertices

1.4.1 Proposition: Counting Number of Walks

Let A, B adjacency matrices of graphs G_A, G_B defined on a common set of n vertices. Then:

$$(AB)_{ij} = \#\text{ of walks from } v_i \text{ to } v_j$$

where for $k \neq i, k \neq j$:

- *we take an edge from G_A (from v_i to v_k)*
- *we take an edge from G_B (from v_k to v_j)*

Similarly, if $k \geq 1$:

$$(A^k)_{ij} = \#\text{ of walks from } v_i \text{ to } v_j \text{ of length } k$$

Proof. This can be proved by induction (see this).

□

1.5 Definition: Connected Vertices

*Let v_i be a vertex in a network $G = (V, E)$. v_i is **connected** to $v_j \in V$ iff there exists a **walk** between v_i, v_j.*

1.6 Definition: Strongly Connected Networks

*A **network** G is **strongly connected** if any pair of vertices v_i, v_j are **connected**.*

1.6.1 Proposition: Strongly Connected from Adjacency Matrix

*Let G be a network with **adjacency matrix** A. The following are **equivalent**:*

1. *G is a **strongly connected** network*
2. $$\forall i, j, i \neq j \ \exists k \in \mathbb{N} \ : \ (A^k)_{ij} > 0$$
3. *A is **irreducible***

Proof.

- ② $\iff$ ③ is immediate, as ② is essentially teh definition of an irreducible matrix.

- ① $\iff$ ② is also immediate, since A^k counts the number of walks of length k, so if $(A^k)_{ij}$ is non-zero for some k, there is a walk between v_i, v_j, so they are connected.

□

1.7 Definition: Clique

*A **clique** is a **network** where each **vertex** is connected to every other **vertex**.*

---

*The **adjacency matrix** of a **clique**, denoted $\mathbf{1}$ contains all 1s, excepts 0s along the main diagonal.*

1.8 Definition: Complementary Network

*Let $G = (V, E)$ be a **network**. A **complementary network** is the network $G' = (V, E')$, whereby E' is the set of all admissible edges not in E.*

---

*If G, G' have **adjacency matrices** A, A', then:*

$$A' = \mathbf{1} - A$$

1.9 Degrees in Networks

1.9.1 Definition: Degree of Vertex

*Let $G = (V, E)$ be a **network**. The **degree** of $v \in V$ is the number of **edges** that connected to v.*

- **How can the degree of a vertex be computed from the adjacency matrix of an undirected network?**
  - compute the ith row/column sum of the **adjacency matrix** A

1.9.2 Definition: Degree Distribution

*The **degree distribution** $P(d)$ represents the **probability** that a random vertex has degree d.*

- **How do typical degree distributions look?**
  - $P(d)$ is typically a long tailed distribution, defined by a power law:

$$P(d) \sim d^{-\gamma}$$

  - normally $\gamma \in [2, 3]$

1.9.3 Proposition: Average Degree of a Network

*If $G = (V, E)$ is an **undirected network**, the **average degree** is given by:*

$$\langle d \rangle = \sum_d dP(d) = \frac{2|E|}{|V|}$$

Proof. This is just the number of edges per node, where we use $2|E|$ since each edge in an undirected network is actually 2 edges.

□

1.9.4 Lemma: Hand Shake Lemma

*Let $G = (V, E)$ be a **network**. Then:*

$$\sum_{v \in V} \deg(v) = 2|E|$$

Proof. Since G is undirected, each edge incides on exactly 2 vertices. Since the degree of a vertex is the number of edges inciding on it, the sum of all degrees must be twice as much as the number of edges. □

1.9.5 Definition: In and Out Degree

*Let $G = (V, E)$ be a **directed network**. Then ,for some vertex v:*

- *the **in-degree** is the number of **edges** incoming to v*
- *the **out-degree** is the number of **edges** outgoing from v*

*Both **in-degrees** and **out-degrees** must sum up to $|E|$.*

1.9.6 Definition: Regular Network

*A **regular network** is a network where **all vertices** have the **same degree**.*

1.9.7 Remark: Friendship Paradox

*The **Friendship Paradox** states that the **average number of friends** of a friend is **smaller** than the **average number of friends** of oneself.*

Proof. The average number of friends of some person corresponds to the average degree of a network, which we know to be:

$$\mu = \frac{2|E|}{|V|}$$

The average number of friends of a friend corresponds to selecting some random vertex (with at least one friend), and then computing its average number of friends. To do this, we can sample a random edge, and then select one of the endpoints. The probability of selecting a vertex v through this strategy is:

$$\frac{\deg(v)}{|E|} \times \frac{1}{2}$$

where $\frac{\deg(v)}{|E|}$ is the probability of picking an edge which contains v, and $\frac{1}{2}$ compensates for the fact that one of two vertices must be chosen.

Then, the average number of friends of a friend is:

$$\nu = \sum_{v \in V} \frac{1}{2} \frac{\deg(v)}{|E|} \deg(v)$$

Now, the variance of degree is:

$$\sigma^2 = \frac{\sum_{v \in V} \deg(v)^2}{|V|} - \mu^2$$

Hence:

$$\nu = \frac{|V|}{2|E|}()\mu^2 + \sigma^2) = \frac{\mu^2 + \sigma^2}{\mu} = \mu + \frac{\sigma^2}{\mu}$$

□

2 Random Graph Models

2.1 Definition: Random Graph

A ***random graph*** *is a sample from a* ***probability distribution*** *over the set of all possible* ***graphs****.*
Equivalently, this can be a distribution over all possible ***adjacency matrices*** *(which for an* ***undirected random graph*** *must be* ***symmetric****,* ***binary*** *and with 0s along the daigonal)*

2.1.1 Proposition: Number of Random Undirected Graphs

Given n vertices, the number of ***undirected graphs*** *is:*

$$2^{n(n-1)/2}$$

Proof. There are:

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

of picking vertices with edges between them. Since matrices are binary, for each entry we have 2 choices as to whether there is an edge or not.

Alternatively, there are $\frac{n(n-1)}{2}$ non-diagonal entries in a $n \times n$ matrix, and each entry has a binary choice.

□

2.2 Definition: Expected Value of a Matrix

*Let A be a **random graph**, the expected value of A is the matrix $\langle A \rangle$ whose entries are:*

$$\langle A \rangle_{ij} = p_{ij}$$

where p_{ij} is the (independent) probability of vertices v_i, v_j sharing an edge.

In particular, any $\langle A \rangle \in S$, where S is the set of matrices which:

- *are **real valued***
- *are **symmetric***
- *have 0s along the **main diagonal***
- *non-diagonal elements have values in $[0, 1]$*

2.3 Proposition: Probability Distribution from Expected Value

*The **probability** of an **adjacency matrix** A is:*

$$P(A) = \prod_{i=1}^{n} \prod_{j=1}^{n} \langle A \rangle_{ij}^{A_{ij}} (1 - \ A_{ij}^{1-A_{ij}})$$

2.4 The Erdös-Rényi Graph

2.4.1 Definition: Erdös-Rényi Graph

*An **Erdös-Rényi Graph** (ERG) (denoted $G(n, p)$) is a **random graph** generated by, for each upper triangular entry, setting the entry to 1 with **iindependent probability** p.*

- **What is the expected value of an ERG?**
  - if $G(n, p)$ is an ERG, then $\langle G \rangle = \mathbf{1}p$

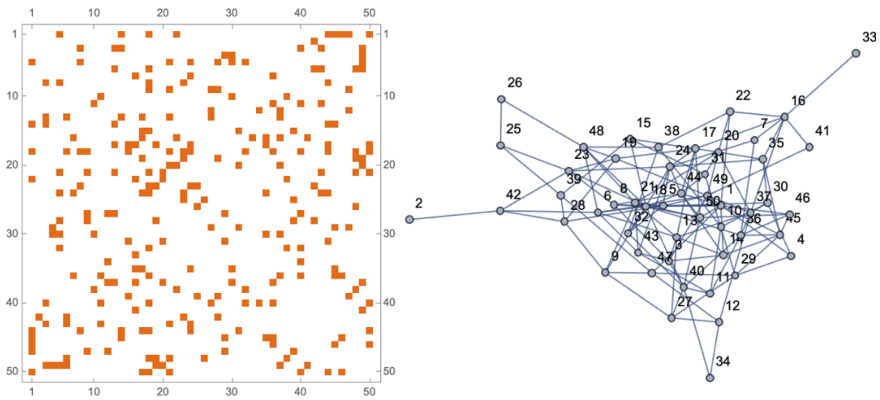


Figure 1: Example of an ERG, both as an adjacency matrix and as a network.

2.4.2 Proposition: Distribution Over Edges and Degree

Let $G(n,q)$ be a ***ERG****. Let m denote the number of* ***links*** *in G, and d a* ***degree*** *in G. Then, we have the following* ***probability distributions****:*

$$p(m) = \binom{n(n-1)/2}{m} q^m (1-q)^{n(n-1)/2-m}$$

$$p(d) = \binom{n-1}{d} q^d (1-q)^{n-1-d}$$

Moreover:

- *the* ***expected number*** *of* ***edges*** *is:*

$$\langle m \rangle = \frac{qn(n-1)}{2}$$

- *the* ***average degree*** *is:*

$$\langle d \rangle = q(n-1)$$

Proof. In ERGs, we are essentially making choices according to a bionomial distribution: we have $\frac{n(n-1)}{2}$ independent events (graphs with n nodes) and a probability q of a success (placing an edge).

For the degree, we have $n-1$ independent events (number of vertices which can be connected to some other vertex), and the probability of success (2 vertices joined by an edge) is q.

The expected values come from the expected value for binomial distributions, whereby if $X \sim \text{Bin}(n,p)$ then $E(X) = np$. □

2.4.3 Proposition: Average Distance Between Vertices

The ***average distance*** *between pairs of vertices in an* ***ERG*** *of n nodes is:*

$$\approx \frac{\log(n)}{\log(\langle d \rangle)} = \frac{\log(n)}{\log(q(n-1))}$$

2.4.4 Remark: Expected Degree in ERGs

The ***average degree*** *in ERGs depends linearly on n, which might be undesirable. Because of this, sometimes we set:*

$$q \propto \frac{1}{n}$$

so that $\langle d \rangle \propto 1$ as $n \to \infty$.

In fact, as $q \to 0$, the binomial distribution over degree approximates a ***Poisson distribution*** *with $\mu = \langle d \rangle$:*

$$p(d) = \frac{\langle d \rangle^d}{k!} e^{-\langle d \rangle}$$

2.5 Definition: Stochastic Block Model

A ***stochastic block model*** *is a generalisation of ERGs, whereby $\langle A \rangle$ contains* ***blocks*** *of probabilities.*
In particular, ***vertices*** *within the* ***same block*** *have some shared probability of conforming edges. There is also fixed probabilities for* ***edges*** *between* ***vertices*** *of* ***different blocks****. Nonetheless,* ***edge probabilities*** *are all* ***independent****.*

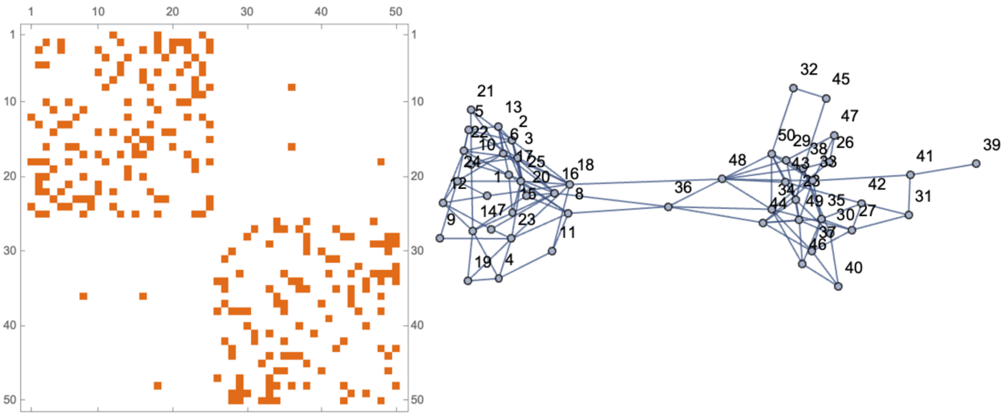


Figure 2: An example of a **stochastic block model**. This could for example be used to model the interaction between people from one same college (vertices), whereby each block corresponds to students in sciences, or students in humanities.

2.6 Definition: Configuration Model

*A **configuration model** is another generalisation of **ERGs**, whereby we construct **random graphs** where each vertex v_i must have some fixed degree d_i.*

---

*To generate these graphs, given some degree sequence $d_1, \dots, d_n$, we create d_i **stubs** (half vertices) for each vertex v_i. Then, we randomly **connect** the stubs, taking care to not form multiple edges or self-loops.*

2.6.1 Proposition: Expected Number of Links in Configuration Models

*Let A be the **adjacency matrix** of a **configuration model** with degree sequence $d_1, \dots, d_n$ and edge set E. Then:*

$$\langle A \rangle_{ij} = \frac{d_i d_j}{2|E|}$$

Proof. By the handshake lemma, we have that:

$$\sum_{i=1}^{n} d_i = 2|E|$$

so the probability of an edge (v_i, v_j) is:

$$\frac{d_j}{2|E|}$$

Hence, the expected number of links between v_i and v_j is given by:

$$\frac{d_i d_j}{2|E|}$$

(since v_i has d_i possible stubs to which v_j can connect)

□

3 Measures Derived from Walks and Paths

3.1 Definition: Distance Between Vertices

*Let $G = (V, E)$. The **distance** between v_i, v_j is:*

$$\delta(v_i, v_j) = \text{smallest number of } \textbf{edges} \text{ in paths between } v_i \text{ and } v_j$$

This can be computed via:

$$\delta(v_i, v_j) = \min_{l \geq 1}\{l \mid (A^l)_{ij} > 0\}$$

- **When does this measure of distance satisfy the definition of a distance?**
  - when we have **undirected networks**, then δ satisfies:
    - **non-negativity**
    - $\delta(v_i, v_j) = 0 \iff v_i = v_j$
    - **symmetry**
    - **triangle inequality**
  - for **directed networks**, **symmetry** doesn't apply

3.2 Proposition: Dijkstra's Algorithm

Dijkstra's Algorithm *is an algorithm used to compute the distance of some vertex v_i to any other vertex v_j of the graph.*

The procedure is as follows:

1. *For some v_i, fix:*
$$\delta(v_i, v_j) = \infty(1 - \delta_{ij})$$
2. *Pick any neighbour v_j of v_i and set:*
$$\delta(v_i, v_j) = 1$$
Declare v_i as visited
3. *For any neighbour v_l of v_j (except v_i), set:*
$$\delta(v_j, v_l) = \min(2(= \delta(v_i, v_j) + 1), \delta(v_i, v_l))$$
4. *Once we've visited all the nighbours of v_j, declare it as visited.*
5. *Select an unvisited vertex with the smallest distance value (2 given the previous iteration)*
6. *For each of the neighbours, we repeat the 3 previous steps , until every vertex has been visited.*

This is a great video explaining Djikstra's Algorithm on a small example graph.

- **For what sort of graphs shoudl Djikstra's Algorithm be used?**
  - when graphs are **large** and **sparse**
  - repeatedly multiplying A to obtain A^k can become innefficient otherwise

3.3 Definition: Average Distance of a Network

*The **average distance** of a **network** is the average distance over all pairs of **distinct vertices**:*

$$L = \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{i-1} \delta(v_i, v_j)$$

- **Is the average distance in real life networks large?**
  - L tends to be small relative to the number of vertices
  - for example, Facebook:
    * $n \approx 7.2 \times 10^8$ active users
    * $\approx 6.9 \times 10^10$ friendship links
    * but $L \approx 4.7$

3.4 Definition: Diameter of a Network

*The **diameter** of a network is the **longest** walk between any vertex pair:*

$$D = \max_{u,v \in V} \delta(u, v)$$

3.5 Definition: Strong and Weak Connectivity in Directed Networks

In undirected networks, connectendess is an equivalence relation (reflexive, symmetric and transitive). However, this no longer makes sense for directed networks, so we need the notion of weak and strong connectivity.

*Let u, v be **vertices** in a network. Then:*

- *u, v are **strongly connected** if there exists a **reciprocal walk** between them*
- *u, v are **weakly connected** if there exists a **walk** between them when we discard directionality*

*Both these notions of **connectedness** form an **equivalence relation**.*

- **What is a strongly connected component?**
  - a **maximum set** of **vertices** in which every **vertex pair** is **strongly connected**

4 Clustering Coefficient & Small Worlds

4.1 Definition: Local Clustering Coefficient

Consider a network G and a vertex v_i. The ***local clustering coefficient*** *of v_i is:*

$$C_i = \frac{\#\ \textit{triangles including}\ v_i}{d_i(d_i)/2} \in [0, 1]$$

A ***triangle*** *is a set of 3* ***utually connected vertices****.*

- **How can the local clustering coefficient be interpreted?**
  - if we think of a **network** as a group of friends, C_i measures how many pairs of friends of i are themselves friends

---

4.2 Definition: Clustering Coefficient

Let $G = (V, E)$ be a ***network****. The* ***clustering coefficient*** *of G is the* ***average local clustering coefficient*** *over all* ***vertices****:*

$$C = \frac{1}{|V|} \sum_{i=1}^{|V|} C_i$$

4.3 Example: Clustering Coefficient of Ring Lattice

- consider a network with n vertices, which are laid out as a circle. 2 vertices have an edge if they are separate by at most k vertices, where:

$$k < \frac{n-1}{2}$$

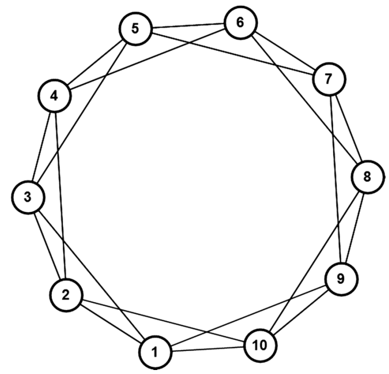
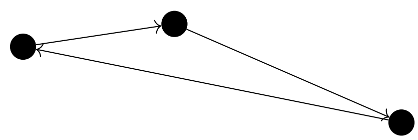


Figure 3: A ring lattice where $k = 2$.

- we know that each **vertex** has **degree** $2k$ (k edges clockwise, k edges anticlockwise) so the denominator for C_i is:

$$\frac{2k(2k-1)}{2} = k(2k-1)$$

- triangles are formed by 2 "clockwise" edges and 1 "anticlockwise" edge (by symmetry the actual direction won't matter)
- moreover, once we choose the 2 edges, the third edge (in the opposite direction) is uniquely defined



- the number of ways of picking 2 clockwise edges (out of all k possibilities) is:

$$\binom{k}{2} = \frac{k(k-1)}{2}$$

- lastly, each vertex can be part of a triangle in 3 different ways (depending on which of the 3 clockwise/anticlockwise edges stem from it), so we get:

$$C_i = \frac{3k(k-1)/2}{k(2k-1)} = \frac{3(k-1)}{2(2k-1)}$$

- by symmetry, every vertex has this local clustering coefficient, so:

$$C = C_i = \frac{3(k-1)}{2(2k-1)}$$

- thus, as $k \to \infty$:

$$C \to \frac{3}{4}$$

  which is fairly large

4.4 Definition: The Small World Model

*The **small world model** is created by modifying the **ring lattice**. In particular, with probability p an **edge** of the lattice is rerouted to some other vertex (chosen uniformly at random).*

---

*A variation of this doesn't destroy **edges**, and simply creates new ones between vertices.*

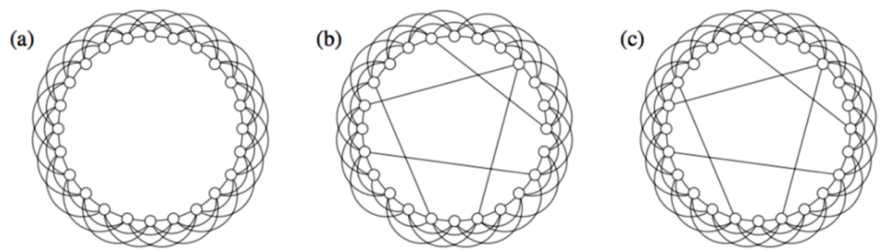


Figure 4: a) Is the standard ring lattice with $k = 3$. b) Is the **small world model** c) Is the variation on the **small world model**

- **What happens to the small world model's clustering coefficient as $p \to 1$?**
  - as $p \to 1$, almost all edges get rerouted
  - this will change the network to become more similar to the ERG
- **What is the purpose of the rewiring process in the small world model?**
  - the standard **ring lattice** has a diameter of around $\frac{n}{2}$ (there are n nodes; the longest walk will occur if you go through adjacent nodes to a node which is diammetrically opposite to you)
  - by rewiring, we reduce the maximum number of steps required to get between nodes
- **Where are small world networks prevalent?**
  - in **social networks**
  - the reduced diameter is exemplified by the phenomenon of "six degrees of separation"

5 The BA Preferential Attachment Model

The BA model is a dynamic network which evolves over time through the mechanism of preferential attachment. We use this as an example of how networks in the real world might eveolve over time.

5.1 Definition: BA Model

*The **BA Model** constructs an **evolving network** using the following steps:*

1. *Consider n_0 initial **vertices**, each with degree at least one (i.e a **clique**)*
2. *Add a new **vertex** to the network, with $m < n_0$ **half-edges**. If the network has n' **vertices** (initially $n' = n_0$) with degrees d_i, the probability that a **half-edge** connects to v_i is given by:*

$$\Pi(d_i) = \frac{d_i}{\sum_{j=1}^{n'} d_j}, \quad i \in [1, n]$$

   *This is the **preferential attachment mechanism**: **vertices** with **higher degree** are more likely to get attached to. However, this must be carried out carefully*[a]

3. *Continue repeating step 2 until we reach a desired number of vertices n.*

[a]During this step, we should avoid generating **multiple edges** between 2 vertices. Moreover, it is a design decision whether we need to update the d_i as new edges are generated through this process

5.2 Proposition: Degree Distribution of the BA Model

*The **degree distribution** of the **BA Model** is given by:*

$$P(d) \propto d^{-3}$$

Proof. We prove this through differential equations, although other methods are possible.

Assume that new vertices are added randomly at an expected rate of 1 per unit time. After t timesteps, we've added t vertices. If we assume the original network is sparse,then there'll be approximately mt edges (since each time we add a vertex we generate m edges) connecting the $t + n_0$ vertices.

Now, let $d_i(t)$ denote the expected degree of the ith added vertex (added at time i). Then, for $t \geq i$:

$$\frac{d}{dt} d_i(t) = \text{rate at which new edges are added} \times P(\text{new edge attaches to } i\text{th vertex})$$

In other words, using the preferential attachment formula alongside the handshake lemma:

$$\frac{d}{dt} d_i(t) = m \times \frac{d_i(t)}{2mt} = \frac{d_i(t)}{2t}$$

We can solve this by separation of variables:

$$\begin{aligned}
&\frac{dd_i}{dt} = \frac{d_i}{2t} \\
\implies &\int \frac{1}{d_i} dd_i = \int \frac{1}{2t} dt \\
\implies &\ln(d_i) = \frac{1}{2}\ln(t) + C \\
\implies &\ln(d_i) = \ln(t^{1/2}) + C \\
\implies &d_i = At^{1/2})
\end{aligned}$$

Using the initial condition:

$$d_i(i) = m$$

it follows that:

$$d_i(t) = m\left(\frac{t}{i}\right)^{1/2}$$

Now, from this we can construct a cumulative distribution, which tells us the proportion of vertices with degree less than some d. Indeed:

$$d_i(t) < d \iff i > \frac{tm^2}{d^2}$$

Hence, at time t there are approximately $n_0 + t - \frac{tm^2}{d^2}$ vertices with degree less than d. This corresponds to there being a proportion:

$$\frac{n_0 + t - \frac{tm^2}{d^2}}{n_0 + t} = 1 - \frac{tm^2}{d^2(n_0 + t)}$$

of vertices with degree less than d. But then, as $t \to \infty$, this proportion tends to:

$$1 - \frac{m^2}{d^2}$$

(provided that $d > m$, since all vertices must have degree greater than or equal to m)

However, this distribution is a cumulative distribution; the density distribution for degrees can be obtained by differentiating, which gives:

$$P(d) \propto d^{-3}$$

as required.

□

5.3 Proposition: Clustering Coefficient of the BA Model

*The **clustering coefficient** for the **BA Model** is approximately:*

$$C \approx \frac{m-1}{8}\frac{(\log(n))^2}{n}$$

which is such that:

$$\lim_{n\to\infty} C = 0$$

- **Why is the BA preferential attachment model unrealistic?**
  - it assumes **new nodes** have acces to information on the **whole network** before deciding how to connect
  - this is clearly not how real world networks evolve
  - however, this can be modified, so that preferential attachment includes **local mechanisms**

6 Centrality

Centrality is a measure of the importance of vertices in a network. One simple such measure is, for example, the degree. We now explore other ways of viewing centrality.

6.1 Definition: Closeness Centrality

*Let $G = (V, E)$ be a **network**. The **closeness centrality** for a **vertex** v_i is the reciprocal of the **mean distance**:*

$$\text{closeness}_i = \frac{|V| - 1}{\sum_{v_j \in V, v_j \neq v_i} \delta(v_i, v_j)}$$

***Closeness centrality** is only well-defined for **connected networks**.*

6.2 Definition: Betweenness Centrality

*Let $G = (V, E)$ be a **network**. The **betweenness centrality** for a **vertex** v_i is the fraction of **shortest paths** passing through v_i:*

$$\text{betweenness}_i = \frac{2}{(n-1)(n-2)} \sum_{j=1, j \neq i}^{n} \sum_{l=1, l \neq i}^{j-1} \frac{\sigma_{jl}^i}{\sigma_{jl}}$$

where:

- *σ_{jl} is the number of **shortest paths** connecting v_j, v_l*
- *σ_{jl}^i is the number of **shortest paths** connecting v_j, v_l which go through v_i*

If $\sigma_{jl} = 0$ (no path between the 2 vertices) then we use $\frac{\sigma_{jl}^i}{\sigma_{jl}} = 0$ in the sum.

6.3 The Katz Centrality

6.3.1 Remark: Weighting Walks

- in real-life applications, **short walks** between vertices may be thought of as "more important" (i.e spreading of infectious diesease - immediate contagion is more important)
- for a given **adjacency matrix**, we can consider the following sum:

$$I + \alpha A + \alpha^2 A^2 + \dots$$

where walks of length k (encoded within A^k) are weighted according to some α^k, where $\alpha \in (0, 1)$

- **if the sum converges**, then we can use the formula for the infinite sum of a geometric series to derive that:

$$(I - \alpha A)^{-1} = I + \alpha A + \alpha^2 A^2 + \dots$$

6.3.2 Definition: Katz Centrality

*Let $G = (V, E)$ be a **network**. The **Katz Centrality** of a vertex v_i is:*

$$\text{Katz}_i = \sum_{j=1}^{|V|} [(I - \alpha A)^{-1}]_{ij}$$

*That is, we take a **weighted sum** of the number of walks starting in v_i, and sum over all **destination vertices**.*

6.3.3 Proposition: Convergence of Katz Centrality

*The **Katz Centrality** is well-defined when:*

$$\alpha < \frac{1}{\rho(A)}$$

*where $\rho(A)$ is the **spectral radiius** of A.*

Proof. The Katz Centrality is undefined if the matrix $I - \alpha A$ is singular (so that $(I - \alpha A)^{-1}$ doesn't exist). In other words,it is undefined whenever:

$$\det(I - \alpha A) = 0 \implies \det\left(A - \frac{1}{\alpha} I\right) = 0$$

where we have used the fact that one can "pull out" scalars from the determinant. Hence, Katz Centrality is undefined whenever $\frac{1}{\alpha}$ is an eigenvalue of A.

Since A is an adjacency matrix, in particular it is positive, so by the Perron-Frobenius Theorem, there is a largest eigenvalue $r = \rho(A)$ which is positive. In particular, if we maintain $1/\alpha$ above r, we are guaranteed to never hit any other eigenvalue. In other words, the Katz Centrality is defined for α satisfying:

$$\frac{1}{\alpha} > r \implies \alpha < \frac{1}{r}$$

□

We are interested in non-zero α, since when $\alpha = 0$, the Katz Centrality just defaults to 1. In other words, the "interesting" values are those for which $\alpha \in \left(0, \frac{1}{r}\right)$.

7 Spectral Properties

7.1 Definition: Normalised Laplacian

*Let A be an **adjacency matrix**. The **Laplacian** of A is constructed by defining a matrix*

$$D = \text{diag}(d_1, \ldots, d_n)$$

*where d_i is the ith row sum of A. The Laplacian of A is the **symmetric** matrix:*

$$L = D - A$$

---

*The **normalised Laplacian** is given by:*

$$\tilde{L} = I - D^{-1/2} A D^{-1/2}$$

where since D is a diagonal matrix:

$$D^{\omega} = \text{diag}(d_1^{\omega}, \ldots, d_n^{\omega})$$

- **What happens to $A, L, \tilde{L}$ when A is regular?**
  - if A is **regular** (all vertices have the same degree), then the 3 matrices have the **same eigenvectors**

7.2 Proposition: Spectral Properties of Laplacian Matrices

*Let A be an $n \times n$ **adjacency matrix**, and let $L, \tilde{L}$ be its **Laplacian** and **Normalised Laplacian** matrices. Then:*

1. *the **eigenvectors** of $L, \tilde{L}$ form an **orthonormal basis***
2. *$L, \tilde{L}$ alwyas have a 0 **eigenvalue**. For L, the corresponding **eigenvector** is:*
$$\underline{u}_1 = (1, \dots, 1)^T$$
*For $\tilde{L}$, the corresponding **eigenvector** is:*
$$\underline{u}_1 = (\sqrt{d}_1, \dots, \sqrt{d_n})^T$$
3. *if A corresponds to an **undirected network** the 0 **eigenvalue** is **isolated** (no other eigenvalues in a neighbourhood)*
4. *if A corresponds to a **connected network**, all non-zero eigenvalues are positive:*
$$0 = \lambda_1 < \lambda_2 \leq \dots \leq \lambda_n$$
5. *The number of **connected components** in G_A is given by the number of **zero eigenvalues** of $L, \tilde{L}$. In particular, a network is **connected** iff $\lambda_2 > 0$. λ_2 is called the **spectral gap**, and its corresponding **eigenvector** is the **Fiedler vector**.*

8 Network Models

8.1 Fitting Models to Data

- **What does it mean to fit a network model to data?**
  - say we observe **entities**,, alongside some evidence of **relationship** between these entities
  - one could think of modelling these relationships as a **network**, and then analysing the data through network theory
  - to **fit** the model to the data would be to come up with **network parameters**, such that the network best represents the observed data
- **What are false positives?**
  - when the **network** has an **edge** which shouldn't be present
- **What are false negatives?**
  - **edges** which aren't present in the **network**, but which shuld be

8.2 Undirected Range Dependent Models

8.2.1 Definition: Range Dependent Random Networks

*A **range dependent random network** consists of an **ordered list** of **vertices** (labelled with $i \in [1, n]$), whereby 2 vertices i, j are **edge-connected** with **independent probability**:*

$$p_{ij} = f(k) \in (0, 1), \quad k = |i - j|$$

*Here, f is a **monotonically decreasing function**, which "forces" vertices with smaller ranges to be more likely connected.*

---

*Notice, by construction $\langle A \rangle$ is a **symmetric Toeplitz matrix** (elements along a given diagonal are all the same).*

A good option for f is:

$$f(k) = \frac{\alpha e^{-\eta k^2}}{1 + \alpha e^{-\eta k^2}}$$

where $\eta, \alpha > 0$.

8.2.2 Proposition: Fitting Range Dependent Random Networks

Say we are given edge data as an adjacency matrix A. Most likely, the vertex list won't be correct. We aim to fit a range dependent random network to the data, by finding an f which is most likely to generate the network given by A.

*Let A be a **binary adjacency matrix** for some data. Then, the **permutation** $\underline{q}$ of the vertex indices which maximises the **likelihood** of A being a **range dependent random network** is obtained by considering the **order** of **increasing elements** of the **Fiedler eigenvector** of A.*

Proof. Let $\underline{q} = (q_1, \dots, q_n) \in \mathbb{N}^n$ denote a permutation of $[1, n]$, representing a possible configuration for the indices of the range dependent random network. Since we assume edge creation is independent, we have that the likelihood of the observed data (the matrix A) given the range dependent random network assumption is:

$$\mathcal{L} = \prod_{i<j} f(|q_i - q_j|)^{A_{ij}} (1 - f(|q_i - q_j|))^{1 - A_{ij}}$$

where:

- $f(|q_i - q_j|)$ is the probability of having observed an edge in A_{ij}
- $1 - f(|q_i - q_j|)$ is the probability of not having observed an edge in A_{ij}

Moreover, notice, we are only iterating ove rth eupper triangular part of A, since A is symmetric.

We can then take logs and manipulate the expression as:

$$\begin{aligned}
\log(\mathcal{L}) &= \sum_{i<j \ \& \ A_{ij}=1} \log[f(|q_i - q_j|)] + \sum_{i<j \ \& \ A_{ij}=0} \log[1 - f(|q_i - q_j|)] \\
&= \sum_{i<j \ \& \ A_{ij}=1} \log[f(|q_i - q_j|)] + \left(\sum_{i<j} \log[1 - f(|q_i - q_j|)] - \sum_{i<j \ \& \ A_{ij}=1} \log[1 - f(|q_i - q_j|)] \right) \\
&= \sum_{i<j \ \& \ A_{ij}=1} \log\left[\frac{f(|q_i - q_j|)}{1 - f(|q_i - q_j|)}\right] + \sum_{i<j} \log[1 - f(|q_i - q_j|)]
\end{aligned}$$

But now, the second term doesn't really depend on $\underline{q}$ (since it is iterating through all the entries in A, and $|q_i - q_j|$) is symmetric. Hence, maximising the likelihood is equivalent to maximising:

$$\log \mathcal{L}' = \sum_{i<j \ \& \ A_{ij}=1} \log\left[\frac{f(|q_i - q_j|)}{1 - f(|q_i - q_j|)}\right]$$

which is a sum of log odds. Now, if we substitute in:

$$f(k) = \frac{\alpha e^{-\eta k^2}}{1 + \alpha e^{-\eta k^2}}$$

It follows that:

$$\begin{aligned}
\log \mathcal{L}' &= \sum_{i<j \ \& \ A_{ij}=1} \log\left[\frac{f(|q_i - q_j|)}{1 - f(|q_i - q_j|)}\right] \\
&= \sum_{i<j \ \& \ A_{ij}=1} \log\left[\frac{\frac{\alpha e^{-\eta |q_i - q_j|^2}}{1+\alpha e^{-\eta k^2}}}{1 - \frac{\alpha e^{-\eta |q_i - q_j|^2}}{1+\alpha e^{-\eta |q_i - q_j|^2}}}\right] \\
&= \sum_{i<j \ \& \ A_{ij}=1} \log\left[\alpha e^{-\eta |q_i - q_j|^2}\right] \\
&\propto \sum_{i<j \ \& \ A_{ij}=1} (q_i - q_j)^2 \\
&= \frac{1}{2} \sum_{i,j=1}^{n} (q_i - q_j)^2 A_{ij}
\end{aligned}$$

But now, recall that the Laplacian of A is positive semi-definite, and:

*Let L be the **Laplacian** matrix of some $n \times n$ matrix A. Then:*

1. $$L\underline{s} = \underline{0}$$

2. *For any $\underline{w} \in \mathbb{R}^n$, we have a **quadratic form**:*

$$\underline{w}^T L \underline{w} = \frac{1}{2} \sum_{i,j=1}^{n} (w_i - w_j)^2 A_{ij}$$

*In other words, L is always **positive-semidefinite** and if $\underline{w}$ is an **eigenvector** of L corresponding to the 0 **eigenvalue**, then the components of $\underline{w}$ must all be equal (so $w_i = w_j$ for any i, j).*

In particular, this means that to maximise the likelihood, we can equivalently minimise:

$$-\log \mathcal{L}' \propto \underline{q}^T L \underline{q}$$

Now, optimising this subject to $\underline{q}$ having integer entries on $[1, n]$ is difficult. Instead, we relax the problem, and optimise when $\underline{q} \in \mathbb{R}^n$. We can then define a permutation, by reordering vertices in increasing order of their corresponding real elements of $\underline{q}$. Note that scalar addition or scalar multiplication don't affect this ordering. In particular, since we only care about the order, we can restrict $\underline{q}$ so that:

$$\|\underline{q}\| = 1 \quad \textit{and} \quad \underline{q} \cdot (1, \ldots, 1)^T = 0$$

In other words, we seek a $\underline{q}$ which is normalised and orthogonal to a vector of just 1s. But recall, this vector of 1s is in fact an eigenvector of L, with eigenvalue 0 (it is in the null space). The eigenvectors of L form an orthonormal basis, so we can pick $\underline{q}$ to be the Fiedler eigenvector (associated with the first (smallest) non-zero eigenvalue).

□

Generally, the eigenvector associated to the smallest non-zero eigenvalue of amatrix will minimise quadratic forms like $x^T Ax$; see this post.

8.3 Directed Stoachastic Block Models