Network science

- analyzing graphs/network
- models
- motifs
- dynamic vs. logical

Network $\rightarrow$ Graph $G = (V, E)$
- directed
- undirected
degree $d(v) = \# \text{ of neighbors}$

in-degree $id(v)$

out-degree $od(v)$

degree distribution

$p(k) \equiv \text{ probability of observing a node of degree } k$

$\mathcal{U} \leftarrow \text{ universe}$

$p(k) = \frac{n_k}{n}$

where $n = |V|$

$n_k = \# \text{ of vertices with degree } k$

$p(1) = \frac{1}{4}$

$p(2) = \frac{2}{4}$

$p(3) = \frac{1}{4}$
Real-world network vs. Random graphs

Power-law degree distribution

Scale-free distribution

\[ P(k) \propto k^{-\gamma} \]

\[ 2 \leq \gamma \leq 3 \]

for most real networks

\[ \log P(k) \propto -\gamma \log k \]

the notion of "mean degree" doesn't convey much,
Average path length \( n_L = \frac{\sum \sum d(u,v)}{n(n-1)} \)

\[ \begin{aligned}
\text{undirected} & \quad \begin{aligned}
&= \frac{2}{n(n-1)} \sum \sum d(u,v) \\
&= \frac{2}{n(n-1)} \sum \sum d(u,v)
\end{aligned} \\
\text{directed} & \quad \begin{aligned}
&= \frac{1}{n(n-1)} \sum \sum d(u,v)
\end{aligned}
\end{aligned} \]

\[ G = (V,E) \quad |V| = n \quad |E| = m \]

Biological networks vs Random graphs

Small-world network
\[ n_L \ll \log n \]

Ultra small-world
\[ n_L \ll \log (\log n) \]

Eccentricity \((u) = \max_{v} \left\{ d(u,v) \right\} \)
Ecc \( (u) = 3 \)

\[
\text{R}(G) = \frac{\text{minimum eccentricity over nodes in } G}{\text{Redius}} = \min_u \{ \text{Ecc}(u) \}
\]

\( R(x) = 2 \)

\( D(G) = \max \text{ eccentricity of any node in } G \)

diameter

\[
\max_{u, v} \{ d(u, v) \}
\]

\( D(G) = 3 \)

Clustering Coefficient

\( C(u) = \frac{\# \text{ edges between neighbors}}{\# \text{ possible edges between neighbors}} \)

degree \( u \geq 2 \)

\[
C(u) = \frac{1}{3}
\]

\[N(u) = \{ v | (u, v) \in E \} \]

\[N(u) = \{ v_1, v_2, v_3 \} \]
\[
\frac{C(G)}{\text{Clustering coefficient of a graph}} = \frac{1}{n} \sum_{u} C(u) \quad \text{average clustering coefficient}
\]

Only for node with \( d \geq 2 \)

Distribution of clustering coefficient
\[
C(k) = \text{clustering coefficient for node with degree } k
\]

Biological networks
\[
C(k) = k^{-\gamma} \quad \text{why?}
\]
Power law clustering coefficient

Random graphs
\[
C(k) = \text{uniformly low}
\]
Almost no clustering

Many small degree vertices have "jars" local clustering
Few high degree nodes with low clustering
Network motif: A subgraph that is frequent in real-data but not in random graphs.

Statistically significant pattern

E. coli, transcription network

$|V| = n = 424$ genes

$|E| = 519$

$p = \frac{1}{n} = \frac{1}{424}$

Null model: Random graph

$p =$ total

$\text{self-edge } p$
\[ |E| \leftarrow \# \text{edges observed in } G^- \]

\[ P(k \text{ self loops}) = \binom{|E|}{k} \cdot p^k (1-p)^{|E|-k} \]

\[ P\text{-value}(40) = \sum_{i=0}^{40} P(k) \leq 1 \times 10^{-10} \]

\[ \Rightarrow \text{highly unlikely event under the null or random model.} \]

\[ \Rightarrow \text{auto-regulation is a network motif} \]

Mean or expected \# of self loops in random graph

\[ \mu = \frac{|E|}{p} \]

\[ \sigma^2 = |E| p (1-p) \]

\[ Z\text{-score}(40) = \frac{40 - \mu}{\sigma} \]

\[ \mu = 1.2 \]

\[ \sigma = \sqrt{\sigma^2} = 1.1 \]

\[ 40 - 1.2 \approx 34.8 \gg 3 \]
Random Graph Model

Erdős–Rényi Random Graphs

- Undirected
- Allow self-loops

Choose one graph at random from the set of graphs with \( n \) nodes and \( m \) edges.

\[
\text{Input} \leftarrow n, m
\]

\[
|V|, |E| = n, m
\]

\[
P(k) = \frac{n(n-1)/2}{M} = \frac{n(n-1) + n}{2} = \frac{n(n+1)}{2}
\]

\[
P(\text{selecting a node of degree } k) = \frac{n(n-1)/2}{M}
\]

for a given node \( u \)
\[ p(k) = \binom{n}{k} p^k (1-p)^{n-k} \]

\[ \text{Expected/mean degree} \]
\[ \mu_d = np \]

\[ \sigma_d^2 = np(1-p) \]

Variance of degree \( \approx np \)

Poisson distribution \( \lambda = np \)

\[ p(k) = \frac{e^{-\lambda} \lambda^k}{k!} \]

Large & Sparse graphs
\( n \rightarrow \infty \) \( p \rightarrow 0 \)

Nothing like a lower law.