Guideline for Macroscopic Structural Characterization and Model Identification

1. Introduction: Generative Models for Complex Networks

Understanding how a network is formed is essential for interpreting its global structure. Here we consider four candidate generative models:

1.1. Erdős–Rényi (ER) Model

In the ER model, every pair of the N nodes is connected with a fixed probability p. Thus, each possible edge exists independently with probability p.

- **Mathematical Notation:**
  The expected degree is given by
  $$\langle k \rangle = p (N-1),$$ and the degree distribution is binomial, which for large N and small p approximates a Poisson distribution: $$P(k) \approx \frac{(\langle k \rangle)^k e^{-\langle k \rangle}}{k!}.$$
- **Expected Characteristics:**
  - Narrow (homogeneous) degree distribution
  - Low clustering coefficient
  - Neutral assortativity (approximately zero)

1.2. Watts–Strogatz (WS) Model

The WS model starts with a regular lattice where each node is connected to k nearest neighbors. Then, each edge is rewired with probability β (with an intermediate value often chosen) to a randomly selected node.

- **Mathematical Notation:**
  Although there is no single closed-form expression, the procedure is as follows:
  1. Create a ring lattice where every node is connected to its k nearest neighbors.
  2. For each edge, rewire it with probability β.
- **Expected Characteristics:**
  - **High clustering coefficient:** Because many local connections are preserved.
  - **Short average path length:** Due to the creation of shortcuts by rewiring.
  - **Narrow degree distribution:** Since most nodes keep approximately k links, with only slight variations from rewiring.

1.3. Barabási–Albert (BA) Model

The BA model is based on preferential attachment, where the network grows by adding nodes one at a time. Each new node connects to m existing nodes with probability proportional to their degree.

- **Mathematical Notation:**
  The probability that a new node connects to node i is

$$\Pi(k_i) = \frac{k_i}{\sum_{j} k_j}.$$ This results in a power-law degree distribution: $$P(k) \sim k^{-\gamma}, \quad \text{with } \gamma \approx 3.$$

- **Expected Characteristics:**
  - **Power-law degree distribution:** A few hubs with very high degree and many nodes with low degree.
  - **Moderate clustering coefficient:** Typically lower than WS but higher than a random graph.
  - **Negative assortativity:** Hubs tend to connect to low-degree nodes.

1.4. Configuration Model (CM)

The configuration model creates a network by first specifying a degree sequence ($\{k_i\}$) (often chosen to follow a power-law) and then randomly connecting nodes while preserving these degrees.

- **Mathematical Notation:**
  If the degree distribution follows a power law, $$P(k) \sim k^{-\gamma},$$ and in our case, if ($\gamma < 2.5$), the tail is heavier than in the BA model.
- **Expected Characteristics:**
  - **Very heavy-tailed degree distribution:** More extreme heterogeneity with the potential for super-hubs.
  - **Low clustering coefficient:** As edges are randomly assigned subject only to the degree sequence.
  - **Potential disassortativity:** Similar to BA, if high-degree nodes preferentially connect with low-degree ones.

---

2. How Each Macroscopic Metric Affects Model Identification

2.1. Number of Nodes ((N)) and Edges ((E))

- **Impact:**
  These values determine the network's overall size and its density, given by
  $$\rho = \frac{2E}{N(N-1)}.$$
- **Model Implications:**
  - **ER Model:** Density is directly controlled by (p).
  - **WS Model:** Starts with a regular structure (high density locally) but becomes sparse as shortcuts are introduced.
  - **BA and CM:** Often yield sparse networks even with heavy-tailed degree distributions.

2.2. Degree Metrics (Minimum, Maximum, and Average Degree)

- **Impact:**
  The range of degrees highlights heterogeneity:
  - **Average degree:** Overall connectivity level.
  - **Maximum degree vs. Average degree:** Indicates presence of hubs.
- **Model Implications:**
  - **ER and WS Models:** Expect a narrow spread (low variance).
  - **BA Model:** High maximum degree relative to the average, due to hubs, with a power-law tail where ($\gamma \approx 3$).
  - **CM:** An even heavier tail (if ($\gamma < 2.5$)), leading to more pronounced hubs.

2.3. Clustering Coefficient

- **Impact:**
  The clustering coefficient measures the tendency of neighbors of a node to be connected.
- **Model Implications:**
  - **ER Model:** Typically low, as edges are formed randomly.
  - **WS Model:** High clustering due to the initial lattice structure.
  - **BA and CM:** Moderate to low clustering; the BA model usually exhibits lower clustering than WS.

2.4. Assortativity

- **Impact:**
  Measures the correlation between the degrees of connected nodes.
- **Model Implications:**
  - **ER and WS Models:** Assortativity is near zero (no strong preference).
  - **BA Model:** Often displays negative assortativity because hubs connect to many low-degree nodes.
  - **CM:** May also show negative assortativity if the imposed degree sequence leads to hub-and-spoke structures.

2.5. Average Path Length and Diameter

- **Impact:**
  These metrics reflect the efficiency of connectivity across the network.
- **Model Implications:**
  - **ER Model:** Short average path length and low diameter (small-world property).
  - **WS Model:** Short average path lengths due to shortcuts, despite high clustering.
  - **BA and CM:** Typically exhibit the small-world property; hubs in BA can greatly reduce path lengths.

2.6. Degree Distribution

- **Impact:**
  The degree distribution reveals how connectivity is distributed among nodes.
- **Model Implications:**
  - **ER Model:** Narrow, bell-shaped (Poisson) distribution.
  - **WS Model:** Also narrow with little variance.
  - **BA Model:** Power-law distribution, linear on a log–log plot, with (\gamma \approx 3).
  - **CM:** Power-law distribution with a heavy tail, especially if (\gamma < 2.5).

---

3. How to Evaluate the Model for a Given Network

Step-by-Step Evaluation

1. **Collect Data:**
   For each network (net1, net2, net3, and net4), compute the following metrics:

   - Number of nodes (N) and edges (E)

- Degree metrics: minimum, maximum, and average degree
- Average clustering coefficient (C)
- Assortativity coefficient (r)
- Average path length and network diameter
- Degree distribution (plot using linear and log–log scales)

2. **Compare with Theoretical Predictions:**

   - **ER Model:**

     - **Expected:** Narrow degree distribution (Poisson-like), low clustering, neutral assortativity.
     - **Inference:** If your network exhibits these features, it is likely an ER model.

   - **WS Model:**

     - **Expected:** High clustering coefficient, short average path length, and a narrow degree distribution.
     - **Inference:** A network with these traits, especially high local clustering, likely follows the WS model.

   - **BA Model:**

     - **Expected:** Power-law degree distribution (visible on a log–log plot) with an exponent (\gamma \approx 3), significant hubs, and slight disassortativity.
     - **Inference:** If the degree distribution is heavy-tailed and hubs are evident, the network likely arises from a BA model.

   - **CM (Configuration Model):**

     - **Expected:** A power-law degree distribution with (\gamma < 2.5), indicating an even heavier tail and extreme heterogeneity.
     - **Inference:** A network with these extreme properties is best described by a configuration model where the degree sequence is imposed.

3. **Draw Conclusions:**

   - Document the values and visualizations for each metric.
   - Compare the observed behavior against the characteristics expected from each model.
   - Select the model that best matches the observed macroscopic structure.

---

4. References

- **Newman, M. E. J. (2010).** *Networks: An Introduction*. Oxford University Press.

- **Watts, D. J., & Strogatz, S. H. (1998).** *Collective dynamics of “small-world” networks*. Nature, 393, 440–442.

- **Barabási, A.-L., & Albert, R. (1999).** *Emergence of scaling in random networks*. Science, 286(5439), 509–512.

- **Costa, L. da F., Rodrigues, F. A., Travieso, G., & Villas Boas, P. R. (2005).** *Characterization of complex networks: A survey of measurements*. Advances in Physics, 56(1), 167–242.

Summary

This guideline provides a structured approach to determine the generative model behind a network by:

1. **Explaining Each Model:**
   - **ER:** $P(k) \approx \frac{(\langle k \rangle)^k e^{-\langle k \rangle}}{k!}$
   - **WS:** Built from a regular lattice with rewiring probability β
   - **BA:** Preferential attachment with $\Pi(k_i) = \frac{k_i}{\sum_j k_j}$ leading to $P(k) \sim k^{-3}$
   - **CM:** Generates a network from a prescribed degree sequence, with $P(k) \sim k^{-\gamma}$ (and if $\gamma < 2.5$, then an extreme heavy tail)
2. **Explaining How Each Metric Affects Model Identification:**
   - Size, degree metrics, clustering, assortativity, and path measures each have characteristic signatures for each model.
3. **Summarizing the Evaluation Process:**
   - Compute all metrics for each network.
   - Compare the observed properties with theoretical expectations.
   - Decide which model (ER, WS, BA, or CM) best fits the network based on the collected data.
4. **Providing References:**
   - Several key texts and articles are cited for further study and confirmation of the theoretical properties.