Introduction to Network Analysis 2021/22 (vi)

Scale-free networks and models, graphs vs networks

You are given three networks in Pajek format (edge list and LNA formats are also available).

- Zachary karate club network (small)
- Map of Darknet from Tor network (medium)
- iMDB actors collaboration network (medium)
- WikiLeaks cable reference network (medium)
- Enron communication network (medium)
- A part of Google web graph (large)

I. Barabási-Albert and Price scale-free graphs

1. **(answer)** Study the following two algorithms for generating Barabási-Albert scale-free graphs $G(n, c)$ and Price scale-free graphs $G(n, c, a)$ using relation $\frac{q+a}{n(c+a)} = \frac{c}{c+a}\frac{q}{nc} + \frac{a}{c+a}\frac{1}{n}$. What is the main difference between the algorithms? What is the time complexity of the algorithms?

```
input   nodes n, degree c
output  undirected scale-free G
  Q ← empty queue
  G ← empty graph
  while not G has n nodes do
      i ← add node to G
      for c times do
          Q.add(i)
          Q.add(j ← Q.random())
          add link between i and j
  return G
```

```
input   nodes n, out-degree c, free a
output  directed scale-free G
  Q ← empty queue
  G ← c isolated nodes
  while not G has n nodes do
      i ← add node to G
      for c times do
          if [0, 1).random() < c/(c + a) then
              Q.add(j ← Q.random())
          else
              Q.add(j ← {0, . . . , i}.random())
          add link from i to j
  return G
```

2. **(code)** Implement both algorithms and generate Barabási-Albert and Price scale-free graphs corresponding to larger networks above. Plot their degree distribution p_k and compute power-law exponents γ of seemingly scale-free distributions using maximum likelihood formula below. Are the results expected or are they surprising?

$$\gamma = 1 + \bar{n}\left[\sum_{i=1}^{n} \ln \frac{k_i}{k_{min} - \frac{1}{2}} \delta(k_i \geq k_{min})\right]^{-1}$$

II. Synthetic random graphs vs real networks

Consider different large-scale properties of real networks. Namely, low average node degree $\langle k \rangle \ll n$, one giant connected component $S \approx 1$, short distances between the nodes $\langle d \rangle \approx \frac{\ln n}{\ln \langle k \rangle}$, high average node clustering coefficient $\langle C \rangle \gg 0$, power-law degree distribution $p_k \sim k^{-\gamma}$, pronounced community structure etc.

1. **(answer)** Design synthetic graph model that generates undirected graphs that are *most different* from real networks.

2. **(code)** Implement generative graph model that *well reproduces* the structure of real undirected networks.

3. **(answer)** Does your model have reasonable interpretation or explanation? Does it also reproduce the structure of real directed networks?