Introduction to Network Analysis 2016/17 (vii)

Small-world & scale-free models, graphs vs networks

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

- A simple toy network for testing (tiny)
- The famous Zachary karate club network (small)
- iMDB actors collaboration network (medium)
- A part of Google web graph (large)

I. Watts-Strogatz small-world networks

1. Study the algorithm for generating Watts-Strogatz small-world networks $G(n, k, p)$ introduced in lectures. Does the algorithm generate networks with realistic structure? What is the time complexity of the algorithm?

2. Try to implement the algorithm and generate Watts-Strogatz small-world networks corresponding to smaller networks above. Compute their average clustering coefficient $\langle C \rangle$ and the average distance between the nodes $\langle d \rangle$. Are the results expected or are they surprising?

II. Barabási-Albert and Price scale-free networks

1. Study the following two algorithms for generating Barabási-Albert scale-free networks $G(n, c)$ and Price scale-free networks $G(n, c, a)$ based on the equality $\frac{q+a}{n(c+a)} = \frac{c}{c+a}\frac{q}{nc} + \frac{a}{c+a}\frac{1}{n}$. 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. Try to implement one of the algorithms and generate Barabási-Albert or Price scale-free networks

corresponding to larger networks above. Plot their degree distributions p_k and compute the power-law exponents γ. of seemingly scale-free distributions. Are the results expected or are they surprising?

III. Synthetic random graphs vs real networks

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

1. Try to design a synthetic graph model that generates undirected graphs that are *most different* from real networks.

2. Try to implement a generative graph model that *best reproduces* the structure of real undirected networks.

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