	Network Science	
Lecture 8: August 11, 2009		Lecturer: Anil Vullikanti
	Scribe: Feng (Andrew) Ge	

Centrality Measure: betweenness, similarity and mixing: homphily, structural balance

8.1 Approximation Algorithm in Calculating Average Clustering Coefficiency

Recall that the *clustering coefficiency* of node v is defined as

$$CC(v) = \frac{\text{\# of nbr pairs, e.g., } (w,\, n) \text{ of } v \text{ that are connected}}{\begin{pmatrix} d(v) \\ 2 \end{pmatrix}}$$

The average Clustering Coefficiency of the whole graph is

$$C_G = \sum_{v=1}^{N} \frac{CC(v)}{N}$$

where N is the total number of vertices.

It is computationally expensive to calculate the aveage Clustering Coefficiency, one approximate algorithm from Dr. Vullikanti's group uses statistics theory to estimate the value by sampling the whole graph verticies. The sequence of this algorithm is:

- Given a graph G, uniformly sample s nodes at random, for example, w_1, w_2, $\cdot$, w_s.
- For each vertex $i = 1$ to s:
  - sample a random pari $(u, v) \in N(w_i)$
  - if $(u, v) \in E$, then $x_i = 1$
- output $X = \frac{\sum_i x_i}{s}$

Such an algorithm only needs the memory complexity of $O(S)$.

Proof:

$$\begin{aligned} E(x_i) = Pr(x_i = 1)(x_i = 1) + Pr(x_i = 0)(x_i = 0) &= Pr(x_i = 1)(x_i = 1) \\ &= \sum_{v \in V} Pr(x_i = 1 | w_i = v) Pr(w_i = v) \\ &= \sum_{v \in V} CC(v) Pr(w_i = v) = \sum_{v \in V} \frac{CC(v)}{N} \end{aligned} \tag{8.1}$$

According to Chebyshev Theorem, we have

$$E(x) = \sum_i \frac{E(x_i)}{s} = C_G \tag{8.2}$$

$$Pr\{|x - E(x)| > \epsilon E(x)\} \leq \frac{var(x)}{\epsilon^2 E(x)^2} = \frac{1}{s \epsilon^2 C_G}$$

In large problems, we don't have enough space and other resources to exhaustively count every vertex, therefore, the above appproximation algorithm is a good way to estimate the average CC.

8.2 Paritioning

8.2.1 Betweenness

To estimate betweenness, we need to

- $\forall$ a vertex pair A and B, send one unit of flow equitably on all shortest path
- $bet(v)$ is the total flow through the vertex v

Procedures:

- $\forall$ node
  - Compute BFS
  - Computer the number of shortest path
  - Compute flows
- Do it on other nodes

8.2.2 centrality measure

Girvan-Newman algorithm is is one of the methods used to detect communities in complex systems.

8.3 Closure

8.3.1 Affiliation Network

People are connected through affiliation of some attributes like location contact and same interest.

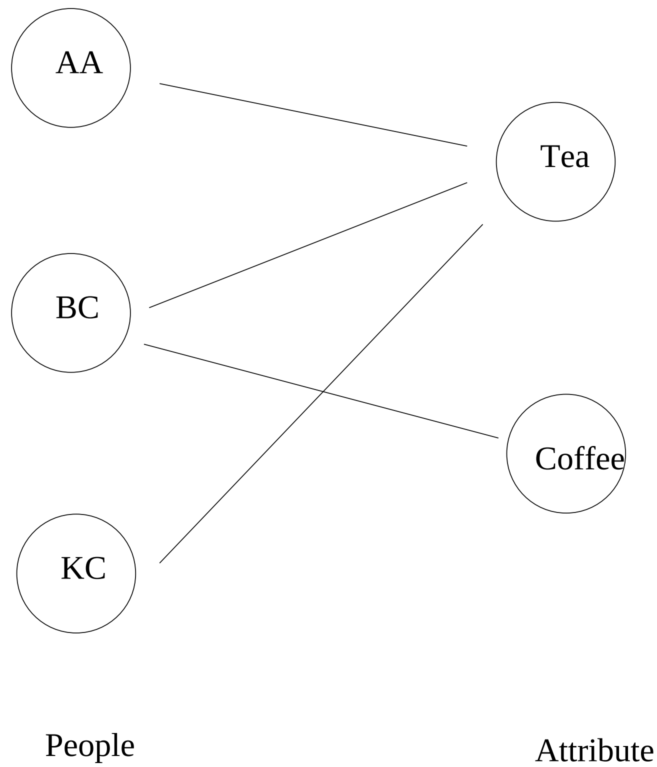

Figure 8.1: One Affiliation Network Example.

8.3.2 Closures

There are a set of closure types: tradic closure, focal closure, and social influenza. They refer to the feature that two nodes might have a strong connection if they are associated with a same neighbor, or a same attribute, or some social influence. If there are multiple such closures, there is a possible amplification effect.

8.3.3 Seggregation Model

Schellieg studied the human mobility in a virtual environment where people select their cell location according to their satisfication which is related to the number of neighbors here.

- People are unhappy if fewer neighbors than t
- Unhappy people switch to a random cell

Here mobility is related to closure, therefore, network structure leads to network dynamics.