Questions and comments CIS6930 presentation pertaining to publication: Mirco Musolesi , Cecilia Mascolo, Designing mobility models based on social network theory, ACM SIGMOBILE Mobile Computing and Communications Review, v.11 n.3, July 2007

For each question, there is response from by Atishay. For few of them, author (Mirco Musolesi) was contacted for comments and his comments appear as "Response from Author"

Question:

The paper used threshold value of 0.25 to define the connectivity matrix, i.e., the edges between the nodes with value < 0.25 are removed before attempts to figure out the communities present in the network. Why a value of 0.25 and what does it imply.

Response from Author: The choice is arbitrary. We had to define a threshold to consider only a subset of "important" relationships (edges) in the network. If you choose a larger threshold you will get a more disconnected graph. We did not investigate the impact of the choice of the threshold on the community detection algorithm itself.

Response from Atishay: My guess is that selection of this threshold may have significant impact on detection of community structure and may alter simulation results. Community Detection is based on algorithms by Newman et al [20] which takes as input a set of nodes and its edges. By choosing different threshold value, we will be altering the input to community detection algorithm and hence altering the result. We are likely to get more or less communities, depending on whether we increase or decrease threshold value, respectively.

Question:

When assigning weights to the edges, paper authors have assigned symmetric weights to the edges. What is the rationale behind assigning symmetric values. It is realistic?

Response from Author: Weights can be asymmetric as you said. We choose to use undirected weighted graphs in order to simplify the model. Please note that for the community detection algorithm we do not use the weights but a threshold-based model. If you use asymmetric weights you need to define a criterion for the threshold (for example you might consider that you keep the link only if max(w_1to2,w_2to1)>threshold or average(w_1to2,w_2to1)>threshold where w_1to2 is the weight representing the value assigned by 1 to the relationship with 2.

Response from Atishay: I understand the point of Author. However, this may be a significant deviation from reality. Only few cases [like where nodes represent a family], may be realistically represented with symmetric edge values. Consider a community of people who are in office, I would find it difficult to argue that the edges should have symmetric values.

Movement of nodes is based on the interaction matrix and that would undergo significant change if we have asymmetric edge values for adjoining nodes.

Response from Author: OK - as I said in my previous email it's only a simplifying assumption.

Question

How the network of nodes & edges is generated in first place and then how communities are detected in it.

Response from Atishay: The network of nodes is generated from a synthetic model called Caveman model [26]. Caveman model is based on research by D J Watts. Caveman model supposedly generates social networks with realistic features like high clustering and low average path length.

I am yet to find full detail of caveman model. Details are not present freely on internet. As per Dr. Mirco, model is described in detail in Watts book Small Worlds The Dynamics of Networks between Order and Randomness. The book is not available in UF for me to read more details. Following is present about the model in the paper and is produced here verbatim.

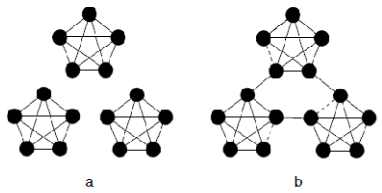

Figure 5: Generation of the social network in input using the Caveman model: (a) initial configuration with 3 disconnected 'caves'. (b) generated social network after the rewiring process.

"The social network is built starting from K fully connected graphs (representing communities living in isolation, like primitive men in caves). According to this model, every edge of the initial network in input is re-wired to point to a node of another cave with a certain probability p. The re-wiring process is used to represent random interconnections between the communities. Figure

5.a shows an initial network configuration composed by 3 disconnected communities (caves) composed by 5 individuals; a possible social network after random rewiring is represented in Figure 5.b."

Once the network is generated, next process is of detecting communities. Paper authors use technique proposed by Newman et al [20]. Here is brief summary of techniques used in the paper.

Weights are immaterial in this process. Once we have the network and its edges, the process figures out which edges are inter-community edges. Consider two communities connected with few inter-community edges, if a inter-community edge is removed, then average distance between two vertices in the network [do not confuse it avg distance b/w 2 communities, communities are not yet known, they are in process of getting detected.] will have maximum increase. In other words, many shortest paths between nodes will pass through these edges.

Such nodes/edges are called with nodes/edges with high centrality.

The process one-by-one detects nodes with high centrality and then one of the edges of that node is removed. The final result is a network which has few communities in it. Now the question is figure out when to stop the process of edge removal?

Solution is based on another parameter introduced in paper by Newman et at [20] called Modularity.

[Following 2 paragraphs taken verbatim from the paper]

"This quantity measures the proportion of the edges in the network that connect vertices within the same community minus the expected value of the same quantity in a network with the same community division but random connections between the vertices.

The algorithm terminates when the obtained value of Q is less than the one obtained in the previous edge removal round."

Question

Is simulation of DSR and AODV protocol performance based on a snapshot?

Response from Atishay: I think the performance is not based on a snapshot. The paper talks about CBR traffic throughout the simulation.

Things learnt from the paper

- Idea about parameters to investigate when studying communities. Which parameters are helpful in comparing synthetically generated traces and real traces!
- We also got to know few properties of real communities like high clustering and low average path length. Similar properties were also introduced in course lectures
- Simulation of protocols should not be on a snapshot and it is not realistic.