Graph Analytics

Kaustubh Kulkarni

Adopted from: Mining of Massive Datasets

Jure Leskovec, Anand Rajaraman, Jeff Ullman
Stanford University

http://www.mmds.org

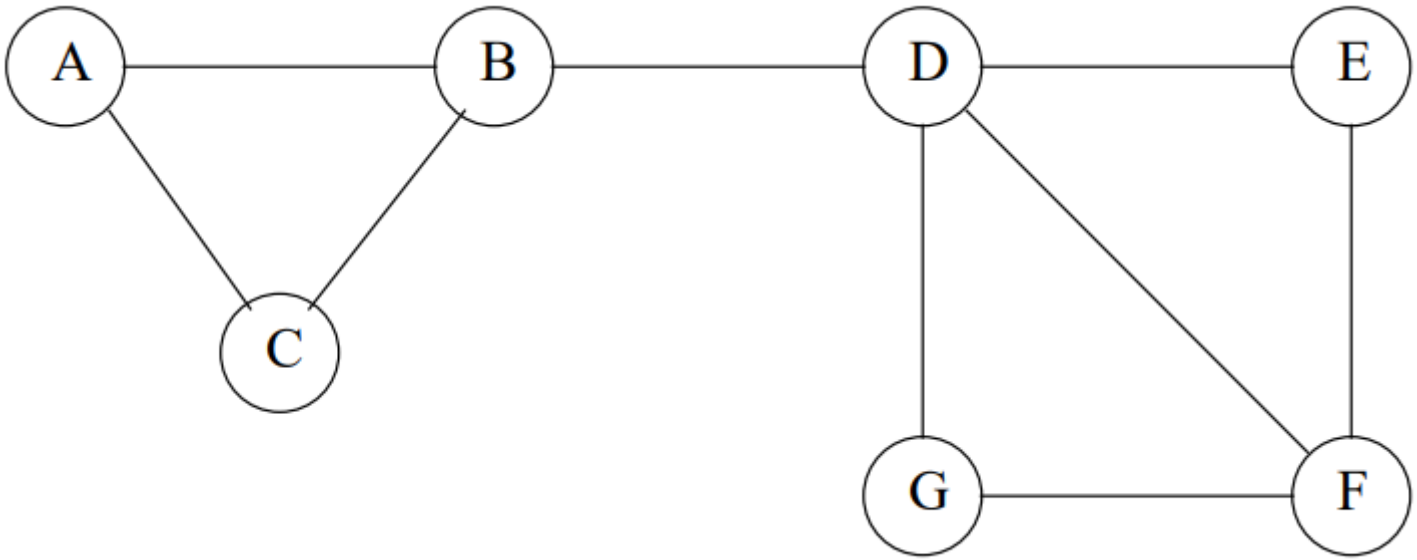
What is a Social Network?

- The essential characteristics of a social network are:
- There is a collection of entities that participate in the network.
- There is at least one relationship between entities of the network
- There is an assumption of non-randomness or locality.

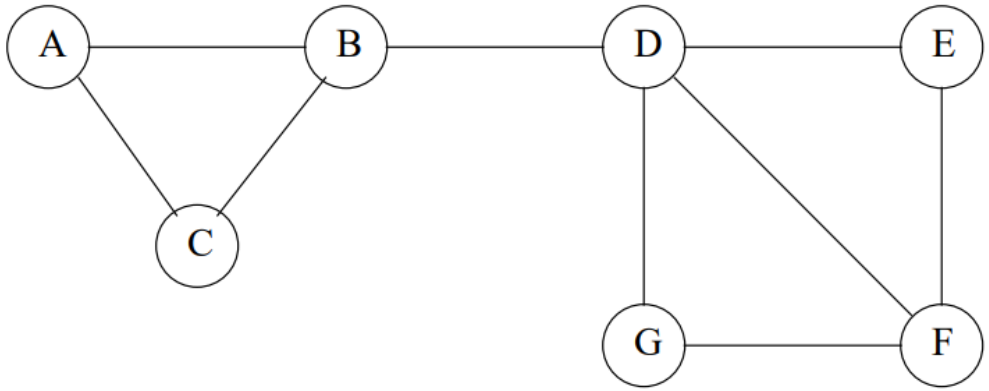
Social Graphs

- Social networks are naturally modeled as graphs, which we sometimes refer to as a **social graph**.
- The entities are the nodes, and an edge connects two nodes if the nodes are related by the relationship that characterizes the network.
- If there is a degree associated with the relationship, this degree is represented by labeling the edges.
- Often, social graphs are undirected, as for the Facebook friends graph.
- But they can be directed graphs, as for example the graphs of followers on X or Instagram.

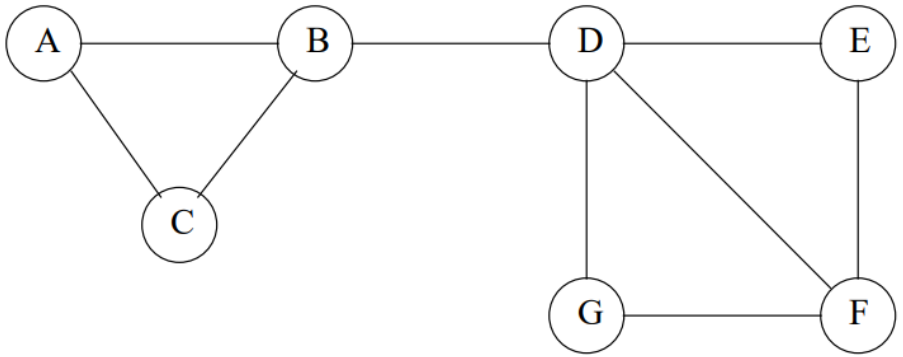
Example of a social graph



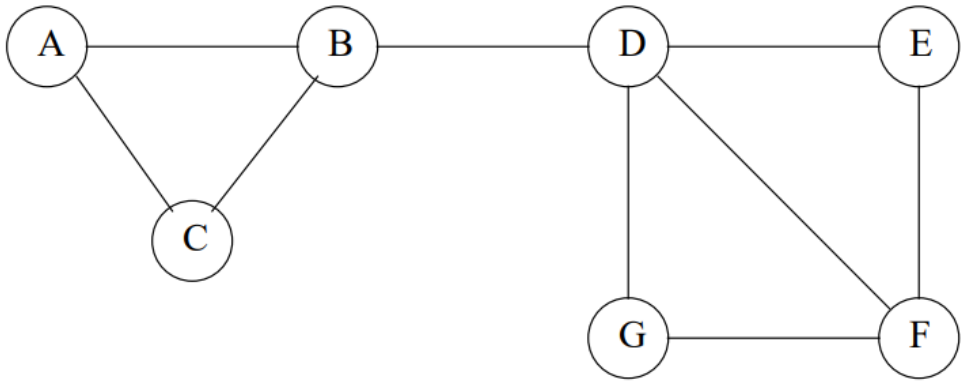
- Figure in the previous slide is an example of a tiny social network.
- The entities are the nodes A through G.
- B is friends with A, C, and D.
- Is this graph really typical of a social network, in the sense that it exhibits locality of relationships?



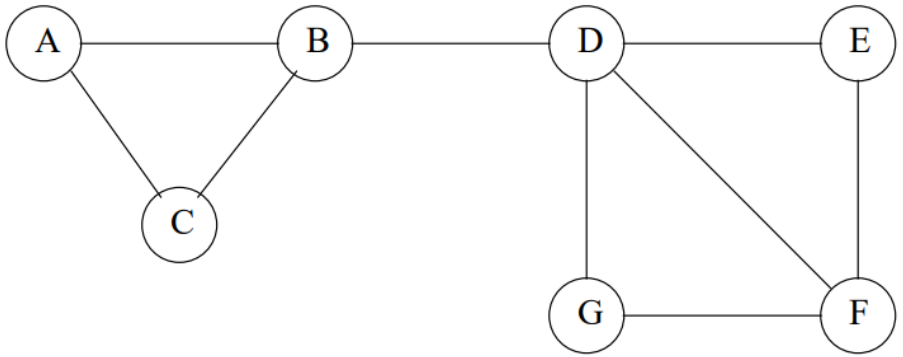
- First, note that the graph has nine edges out of the 7C2= 21 pairs of nodes that could have had an edge between them.
- Suppose X, Y , and Z are nodes, with edges between X and Y and also between X and Z.
- What would we expect the probability of an edge between Y and Z to be?



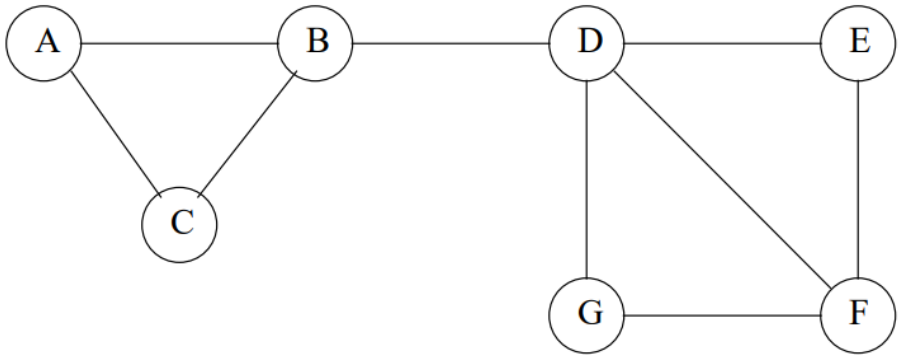
- Since we already know there are edges (X, Y) and (X, Z), there are only seven edges remaining.
- Those seven edges could run between any of the 19 remaining pairs of nodes.
- Thus, the probability of an edge (Y, Z) is 7/19 = .368



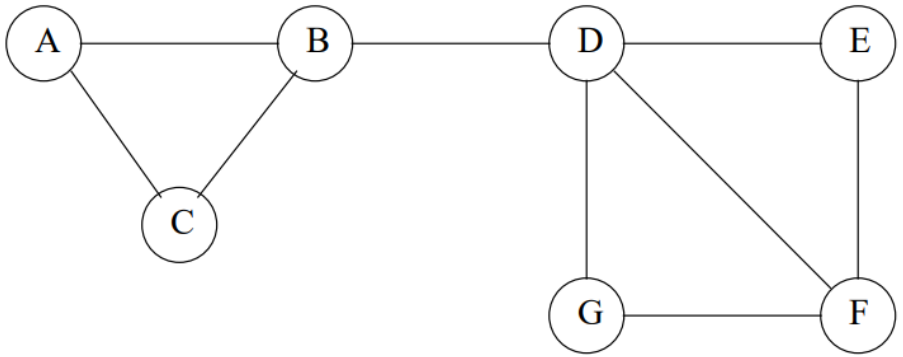
- Now, we must compute the probability that the edge (Y, Z) exists given that edges (X, Y) and (X, Z) exist.
- What we shall count is pairs of nodes that could be Y and Z, without worrying about which node is Y and which is Z



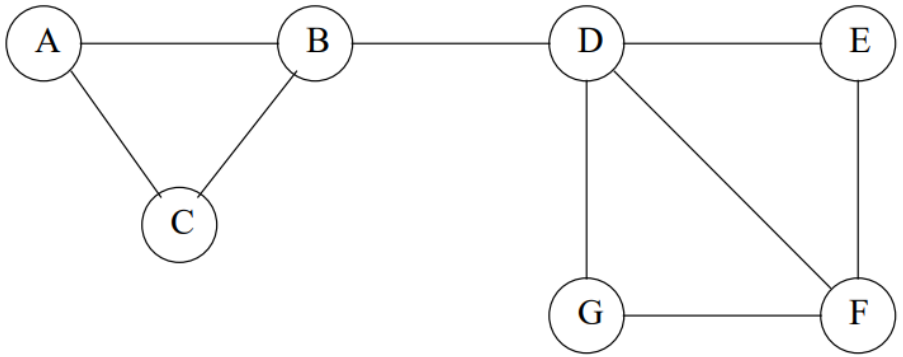
- If X is A, then Y and Z must be B and C, in some order.
- Since the edge (B, C) exists, X=A contributes one positive example (where the edge does exist) and no negative examples (where the edge is absent).



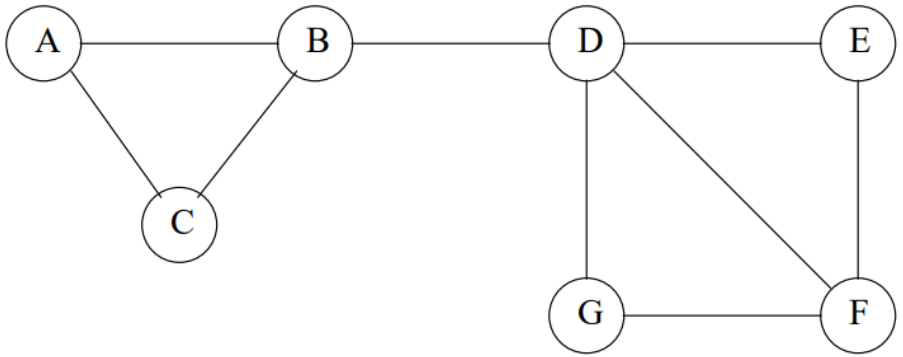
- The cases where X is C, E, or G are essentially the same.
- In each case, X has only two neighbors, and the edge between the neighbors exists.
- Thus, we have seen four positive examples and zero negative examples so far.



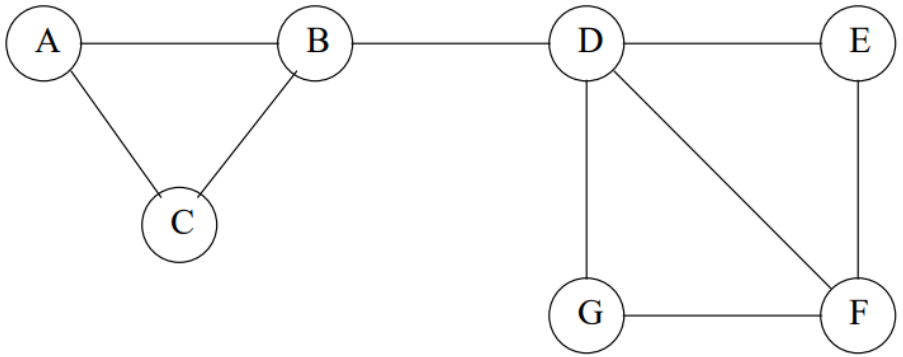
- Now, consider X = F.
- F has three neighbors, D, E, and G.
- There are edges between two of the three pairs of neighbors, but no edge between G and E.
- Thus, we see two more positive examples and we see our first negative example.



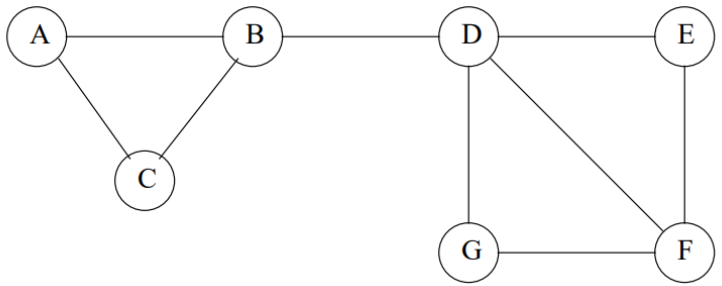
- If X = B, there are again three neighbors, A,C,D.
- But only one pair of neighbors, A and C, has an edge.
- Thus, we have two more negative examples, and one positive example.



- Finally, when X = D, there are four neighbors.
- Of the six pairs of neighbors, only two have edges between them.
- Negative += 4, Positive += 2

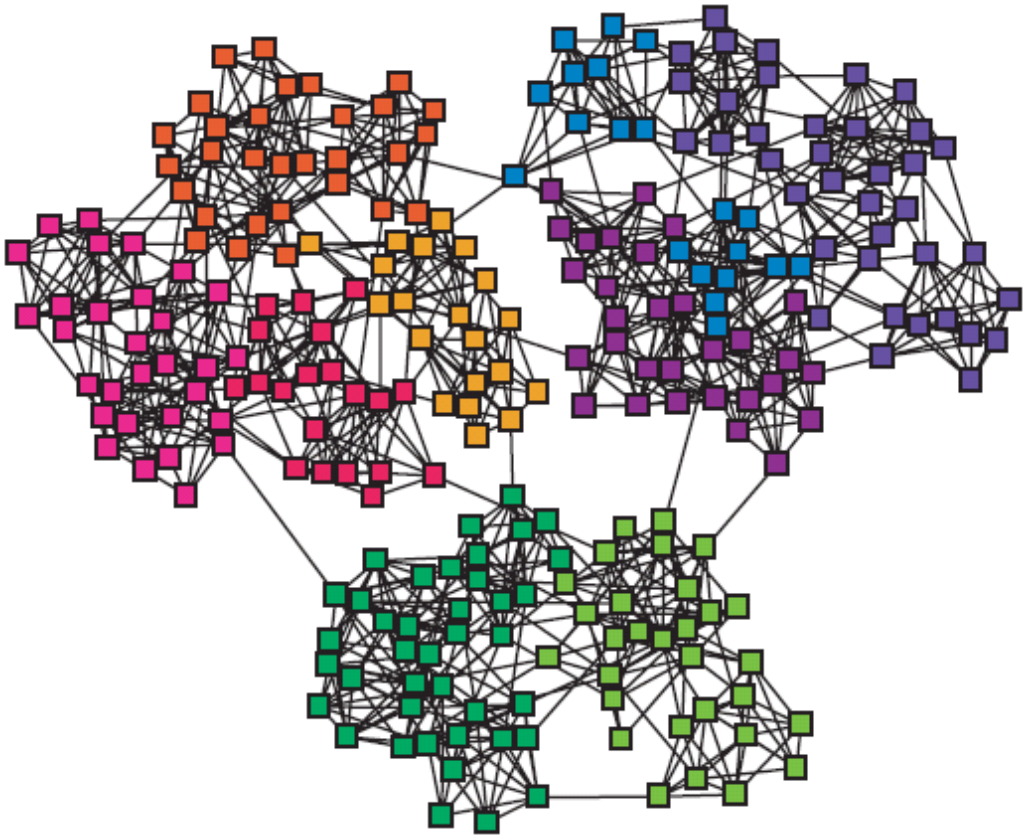


- Thus, the total number of positive examples is nine and the total number of negative examples is seven.
- Positive / (Negative+ Positive) =9/16=0.563
- This fraction is considerably greater than the .368 expected value
- We conclude that our network does indeed exhibit the locality expected in a social network.

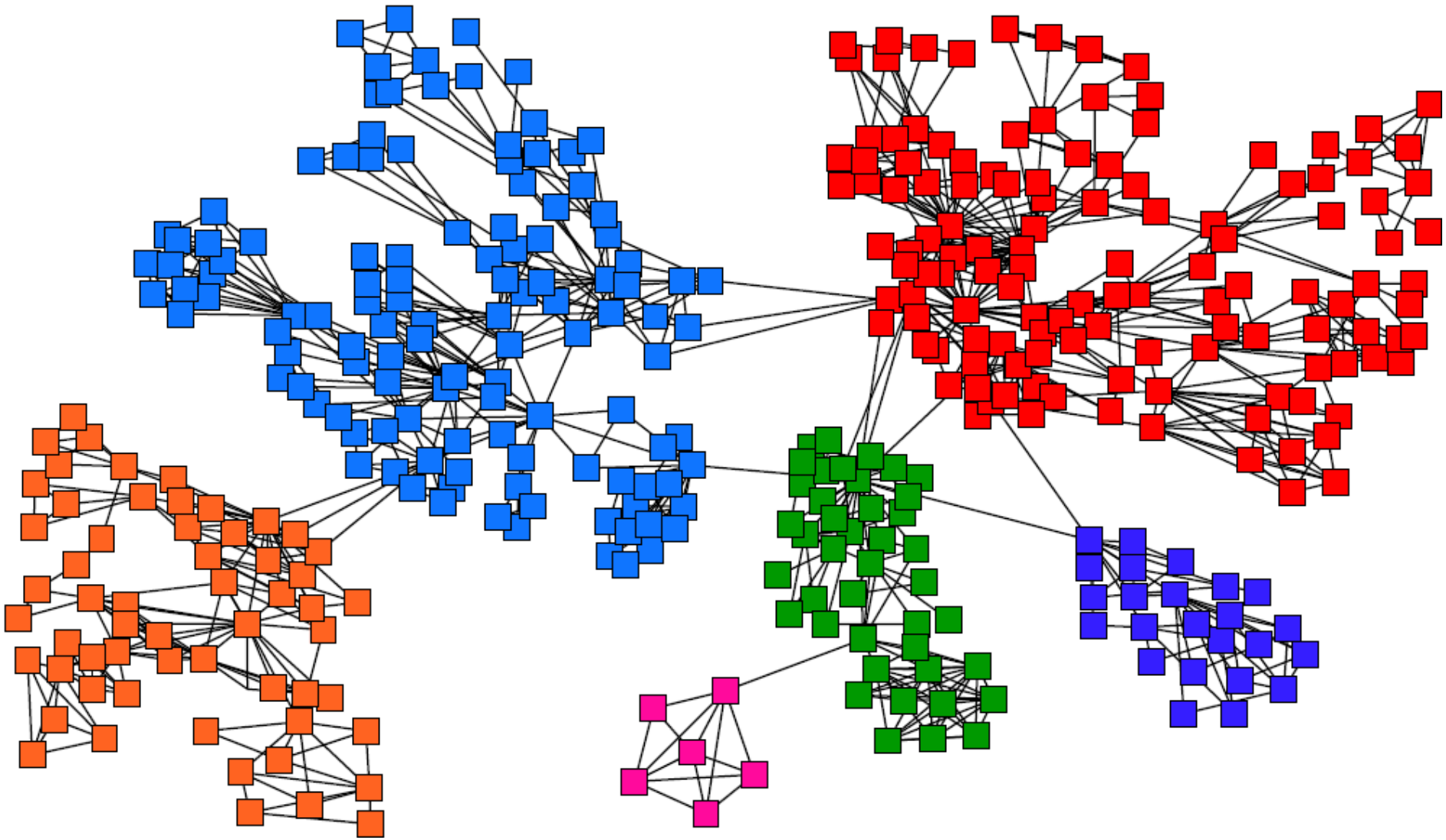


Networks & Communities

- **We often think of networks being organized into modules, cluster, communities:**

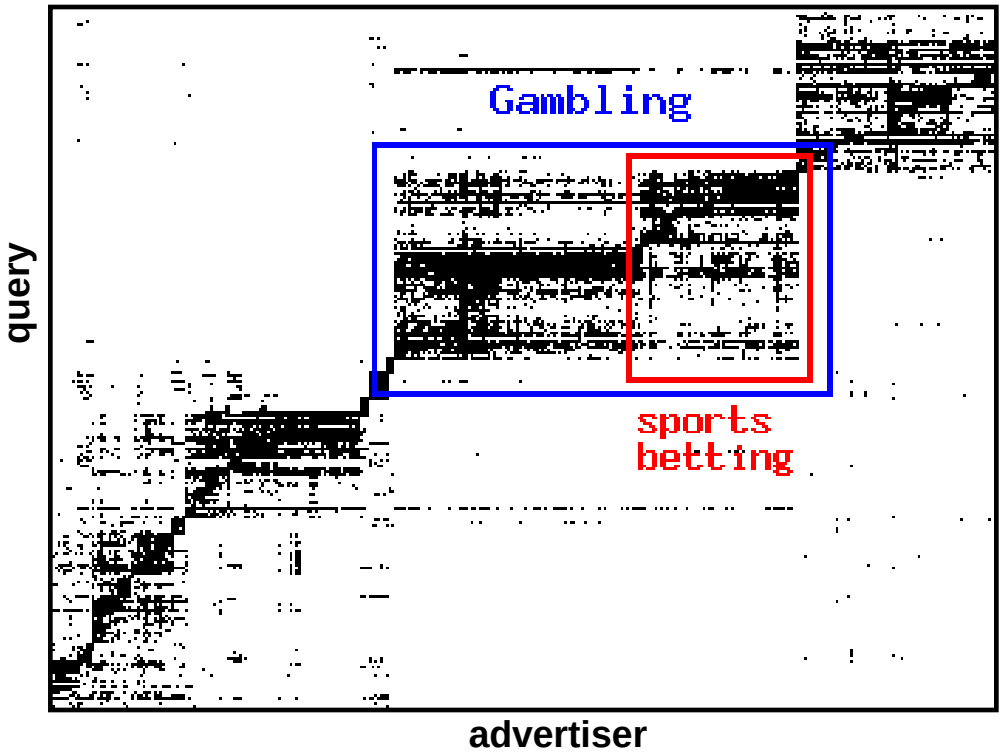


Goal: Find Densely Linked Clusters



Micro-Markets in Sponsored Search

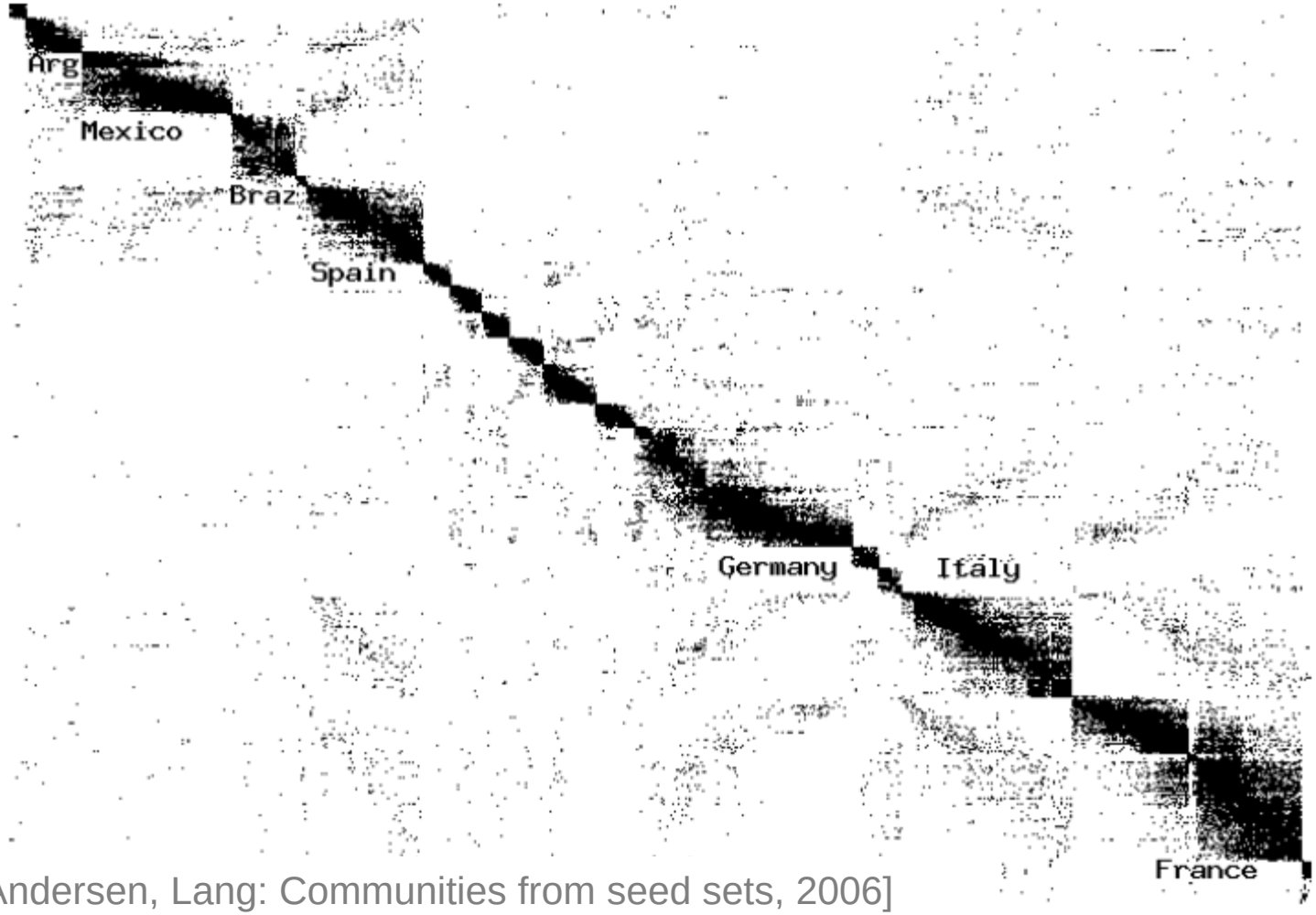
- **Find micro-markets by partitioning the query-to-advertiser graph:**



[Andersen, Lang: Communities from seed sets, 2006]

Movies and Actors

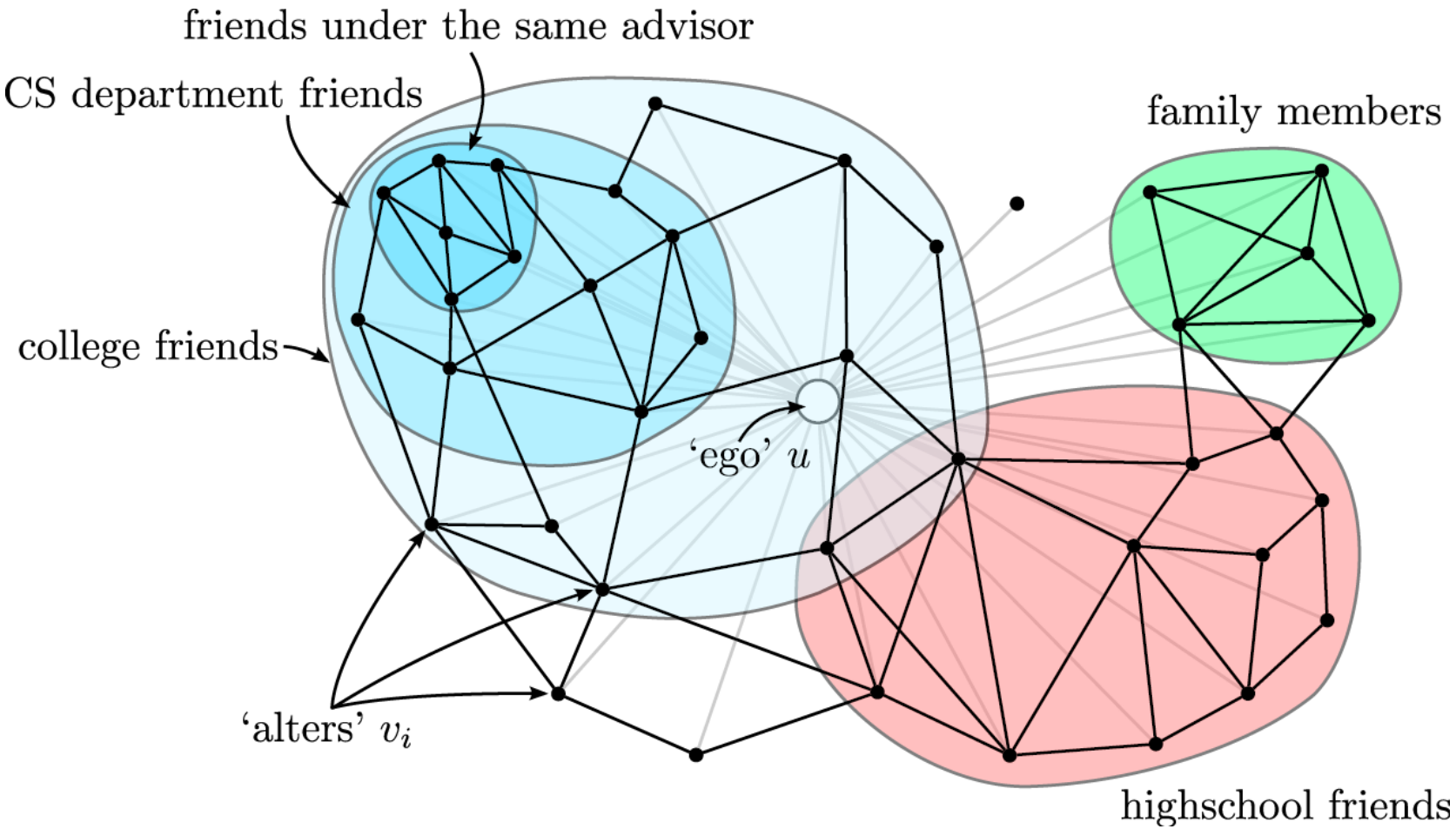
- **Clusters in Movies-to-Actors graph:**



[Andersen, Lang: Communities from seed sets, 2006]

Twitter & Facebook

- **Discovering social circles, circles of trust:**



[McAuley, Leskovec: Discovering social circles in ego networks, 2012]

Questions?