N O T E S
#2 - #5

B. MISHRA

FEBRUARY 20, 2012

Structure and Function

In the last lecture, we argued that we will need to focus on two aspects of a social network: the static combinatorial structure, encoded in a **Graph** and a dynamic strategizing structure, encode in a **Game**. The nodes in the graph are the social agents, constrained to interact in a pairwise manner (possibly directed). Each social agent has now a set of well-defined *neighbors* as well as a *strategy space*, which will be assumed not to be private – in fact, they will be commonly known. We will explore why and how some of these simplifying assumptions will need to be changed. What we will be interested in are:

Who are my friends? Who should I connect to? If everyone connects to everyone else using certain rules, will the social network benefit me? How long will it take for it to become beneficial? Will it remain beneficial in the long-run?

Are my friends reasonable? Will my friends and their friends be rational? Will everyone know that their friends are rational? Will everyone know that all their friends know that all their friends are rational? $\cdots$ Will it be common knowledge that everyone is rational?

Will we do the right thing? How will we know what to do? How should we exercise our strategic choices? How do we know what's good for us? How do we know what's good for all?

Will we get spammed? If my friend is a "mutant," will my world fall apart? Will my friend ever deviate from what should be her best response to my strategic choice?

Will we get exploited? Information asymmetry. If someone has more information than rest of us, can that pose a problem for me?

How soon can we arrive at a "nice" functional social-network?

Graph Theory

Note that the structure of the graph underlying a Social Network will determine its behavior: Think of a graph that is all singletons vs. a completely connected graph. Another example would be a star-like graph representing a market-maker connected to buyers and sellers. In each case, the behavior of the social network will vary widely. The right mathematical framework, in which we can think about this comes from *Graph Theory*.

Graph theory is fundamental to a number of applied fields, including operations research, computer science, and social network analysis. Here, we discuss the basic concepts of graph theory from the point of view of social network analysis.

Graphs: The fundamental concept of graph theory is the *graph*. We start by describing an *(undirected) graph*, a simple (combinatorial) mathematical object describing (irreflexive, symmetric) binary relations on a discrete set.

The binary relation we will describe first will represent the *friendship relation*: It is irreflexive: one is not considered one's friend; It is symmetric: one is friend to a friend; It is usually non-transitive: A friend's friend is not necessarily a friend.

Graphs have a very natural graphical representation (see figure 1), hence the name.

A graph – usually denoted $G(V,E)$ or $G = (V,E)$ – consists of set of *vertices* V together with a set of *edges* $E \subseteq V \times V$. Vertices are also known as *nodes*, *points* and (in social networks) as *actors*, *agents* or *players*. Edges are also known as *lines* or *connection* and (in social networks) as *ties* or *links*. An edge $e = (u,v)$ is defined by the unordered pair of vertices that serve as its end points.

Two vertices u and v are *adjacent* if there exists an edge (u,v) that connects them. An edge $e = (u,u)$ that links a vertex to itself is known as a *self-loop* or *reflexive tie*. The number of vertices in a graph is usually denoted $|V| = n$ while the number of edges is usually denoted $|E| = m$.

$$m \leq \binom{n}{2} = \frac{n(n-1)}{2}.$$

$$\begin{aligned} n &= \text{\# of ways to choose one end of the edge } u \\ (n-1) &= \text{\# of ways to choose the other end of the edge } v \\ & \quad \text{Identify the two ways the same edge can} \\ & \qquad \text{be represented, namely } (u,v) \equiv (v,u). \quad \square \end{aligned}$$

As an example, we can draw a graph (as in figure 1), which has vertex set $V = \{a, b, c, d, e, f\}$ and edge set $E = \{(a, b), (b, c), (c, d), (c, e), (d, e), (e, f)\}$.

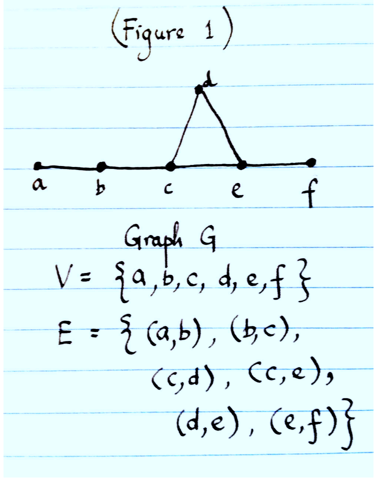
All our graphs will be assumed to be *strict graphs*. We will allow neither *self-loops*:

$$\forall_u (u, u) \notin E;$$

nor *multi-edges*:

$$\forall_{e_1 \neq e_2, e_1 = (u_1, v_1), e_2 = (u_2, v_2)}$$
$$(u_1 = u_2) \rightarrow (v_1 \neq v_2) \wedge (u_1 = v_2) \rightarrow (v_1 \neq u_2).$$

To represent social networks, each line typically represents instances of the *same social relation*, so that if (a, b) indicates a friendship between the person located at node a and the person located at node b, then (d, e) indicates a friendship between d and e. Thus, *each distinct social relation that is empirically measured on the same group of people is represented by separate graphs*, which can have drastically different structures: people you are genetically related to are not necessarily your friends – and *vice versa*.

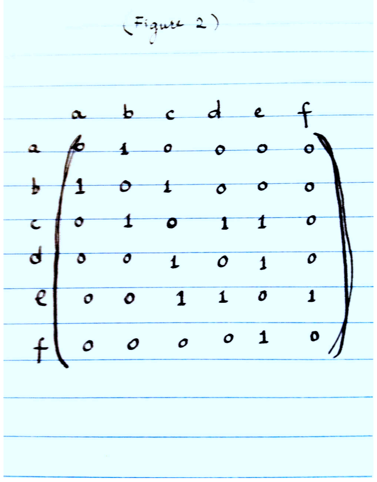


V = {a, b, c, d, e, f}
E = {(a,b), (b,c), (c,d), (c,e), (d,e), (e,f)}

Figure 1: Example Graph

Note that the only information contained in a diagram depicting a graph is adjacency; the position of nodes in the plane (and therefore the length of lines) is arbitrary (unless the underlying geometry has some information). Thus the spatial position of the nodes is completely irrelevant. For example, nodes near the geometric centroid of a graph-diagram are not necessarily more important than nodes on the peripheries.

Every graph has associated with it a (symmetric) *adjacency matrix*, which is a binary $n \times n$ matrix A in which $a_{ij} = 1$ (and $a_{ji} = 1$) iff vertex v_i is adjacent to vertex v_j, [otherwise, $a_{ij} = 0$ and $a_{ji} = 0$]. The natural graphical representation of an adjacency matrix is a table.


(Figure 2)

$$\begin{array}{c|cccccc} & a & b & c & d & e & f \\ \hline a & 0 & 1 & 0 & 0 & 0 & 0 \\ b & 1 & 0 & 1 & 0 & 0 & 0 \\ c & 0 & 1 & 0 & 1 & 1 & 0 \\ d & 0 & 0 & 1 & 0 & 1 & 0 \\ e & 0 & 0 & 1 & 1 & 0 & 1 \\ f & 0 & 0 & 0 & 0 & 1 & 0 \end{array}$$

Figure 2: Adjacency matrix for graph.

Examining the preceding example, we see that not every vertex is adjacent to every other.

Complete Graphs: A graph in which all vertices are adjacent to all others is said to be *complete*. The extent to which a graph is complete is indicated by its *density* (or, equivalently its *sparsity*), which is defined as the number of edges divided by the number of possible total.

If self-loops are excluded, then the number of possible total is

$$\binom{n}{2} = \frac{n(n-1)}{2}.$$

(If self-loops are allowed, then the number of possible total is

$$\binom{n}{2} + \binom{n}{1} = \frac{n(n+1)}{2}.)$$

$$\begin{aligned} \text{density} &= \frac{|E|}{\binom{n}{2}} = \frac{2m}{n(n-1)} \\ &\ \text{(in strict-graphs)} \\ &= \frac{2m}{n(n+1)} \\ &\ \text{(in graphs with self-loops, but without multi-edges)} \end{aligned}$$

Hence the density of our example graph in figure 1 is $6/15 = 0.40$. □

While not every vertex in the example-graph is adjacent, one can construct a sequence of adjacent vertices from any vertex to any other. Graphs with this property are called *connected*. Similarly, any pair of vertices in which one vertex can reach the other via a sequence of adjacent vertices is called *reachable*. If we determine reachability for every pair of vertices, we can construct a reachability matrix R. The matrix R can be thought of as the result of applying *transitive* closure to the adjacency matrix A. Note that A represents a constrained reachability relation, where we only consider the pair of nodes that are reachable from each other in *one or fewer steps*. Similarly A^2 (with addition $+$ interpreted as logical OR: $\vee$ and multiplication interpreted as logical AND: $\wedge$) represents reachability relation, constrained to *two or fewer steps*. In general, A^k (with the appropriate algebraic interpretation as before) is the reachability constrained to *k or fewer steps*. Note that since if a pair of vertices are reachable, then one can be reached from the other in *n or fewer steps* and thus

$$R = A^n.$$

Another way to say this, would be

$$R = A^* = I + A + A^2 + \cdots = A^j|_{\min j : A^j = A^{j+1}}.$$

The parameter $\Delta = \min j : A^j = A^{j+1}$ is rather important in social networks, and is called the *diameter* (or *degrees-of-sepration*) of the graph. If a graph with n vertices has a diameter $\approx O(\lg n)$, then we may call it a "*small-world graph*." We believe that humans have a small-world-graph with about six-degrees-of-sepration, which has apparently gone down to 4.5, with the advent of social networks like `facebook`, `\google+`, etc.

§§§

Degree The number of vertices adjacent to a given vertex is called the *degree* of the vertex and is denoted $d(v)$.

$$\begin{aligned} d(v) &= |\{w|(v,w) \in E\}| \\ &= |\{u|(u,v) \in E\} \end{aligned}$$

It can be obtained from the adjacency matrix of a graph by simply computing each row sum. For example, the degree of vertex c in the graph of figure 1 is 3. The average degree, $\bar{d}$, of all vertices of the graph of figure 1 is 2. In figure 3, vertex c has a degree 4, and the graph's average degree is 2.29.

The minimum degree of a graph G is denoted $\delta(G)$. A vertex with degree 0 is known as an *isolate or a singleton* (and constitutes a component of size 1), while a vertex with degree 1 is a *pendant*.

Note

$$\sum_{v \in V} d(v) = 2|E| = 2m.$$

(Suppose every agent in a social network broadcasts a message to all its friends. Then the total number of messages is just the sum of the degrees of the vertices: $\sum_{v \in V} d(v)$. But for every friendship relation there are two messages traveling between two friends it connects. The total number of messages is twice the total number of friendships: $2|E|$. These two numbers must be same.)

Thus the average degree of a graph is

$$\bar{d} = \frac{\sum_{v \in V} d(v)}{|V|} = \frac{2m}{n}.$$

There is a direct relationship between the average degree, $\bar{d}$, of all vertices in a graph and the graph's density:

$$\text{density} = \frac{|E|}{\binom{n}{2}} = \frac{n\bar{d}/2}{n(n-1)/2} = \bar{d}/n - 1.$$

If a social network has many "well-connected people," then the network is "dense." Thus

$$\bar{d} = \text{density}(n-1). \ \Box$$

Note that if I connect to a well-connected individual in a social-network, then I'll belong to a subnetwork with a high-density. Holding average degree constant, there is a tendency for graphs that contain some nodes of high degree (i.e., high variance in degree) to have shorter distances than graphs with lower variance, with the high degree nodes serving as "shortcuts" across the network. These high-degree nodes are said to be "hubby" and many social network graphs have a fat-tail distribution with an unusually large number of hubby nodes. Kevin Bacon in Hollywood and Paul Erdös in mathematics are supposed to have been examples of this phenomenon. We will try to identify such high density subnetworks, in the forms of *cliques*, *clubs* and *clans*.

Assume that we are in a social network whose average degree is $\bar{d} > 1$. Now, suppose I start with you and start counting all of your friends (who are separated by a DOS (degree of separation) of $= 1$), which on the average is $\bar{d}$. Assuming that they have also $\bar{d}$ friends each (on the average), the total number of individuals separated from you by a DOS ≤ 2 is less than

$$1 + \bar{d} + \bar{d}^2 = \frac{\bar{d}^3 - 1}{\bar{d} - 1}.$$

And so on. Thus the total number of individuals separated from you by a DOS $\leq \Delta =$ the diameter of the graph, will have everyone in the social network (which is assumed to be connected for now): thus,

$$n \leq \frac{\bar{d}^\Delta - 1}{\bar{d} - 1},$$

Or

$$\Delta \lg \bar{d} \geq \lg \left(n(\bar{d} - 1) + 1 \right),$$

Or

$$\Delta \geq \frac{\lg \left(n(\bar{d} - 1) + 1 \right)}{\lg \bar{d}},$$

Thus, in a graph $\bar{d} \geq 2$, $\Delta = \Omega(\lg n / \lg \bar{d})$ gives us a lower-bound. In a Poisson (Erdös-Renyi) random graphs, for large n, average distance can be approximated by $\bar{\Delta} = \lg n / \lg \bar{d}$, where $\bar{d} =$ the expected degree of a node. It can be even shorter, if the graph has some other special properties: "small-worldness."

The notion of a "small world" comes from a Hungarian poet Frigyes Karinthy. In a play, he wrote, he conjectured that any two people among the one and half billion inhabitants of the earth then were linked through at most five acquaintances. Thus by the formula above, we can calculate $\bar{d}$ of such a graph to estimate how many friends one had, on the average, in that world.

$$\exp \left[\frac{\ln 1.5 \cdot 10^6}{5} \right] \approx 68.5.$$

Thus, we can state the above estimation, in terms of a *Karinthy Conjecture*: *Each person should have had approximately 68 "independent" friends.*

In another study, socilogist Stanley Milgram tried to study the "small-world phenomenon" more systematically, and also coined the phrase: "six degrees of separation," in a famous (but somewhat discredited paper) "The Small World Problem." Milgram polled residents of Omaha and Wichita, and asked 160 of them to send a folder to a target recipient by sending it to an acquaintance, who would then do likewise, etc. Forty-two out of hundred and sixty folders found their way to the intended recipients, with a median number of intermediate acquaintances $= 5.5$, which he rounded up to six. Milgram's "six degrees of separation" conjecture suggests that each person has approximately forty-one friends.

§§§

Given what we have understood so far, we would prefer a social network that has high density (at least locally), some highly-connected ("hubby") individuals and some intimate social groups ("cliques") that one can join.

Subgraph: A *subgraph* of a graph G is a graph whose vertices and edges are contained in G. If $H \subset G$ is a subset of G, the subgraph induced by H in G consists of all vertices of H together with all edges of G, which connect vertices $u, v \in H$ in G.

$$\begin{aligned} G &= (V, E) \\ H &= (W, \{(u,v) \in E | u \in W \wedge v \in W\}) \\ &\quad \text{where } W \subseteq V. \end{aligned}$$

Maximal Subgraphs (with respect to a property P): Subgraphs of G satisfying P such that no larger subgraphs with property P exist in G, which contain them.

Clique: A *clique* is a *maximal complete subgraph* K_p in G.

Thus a complete subgraph of G is a section of G that is complete (i.e., has density $= 1$). A maximal complete subgraph is a subgraph of G that is complete and is maximal in the sense that no other node of G could be added to the subgraph without losing the *completeness property*. In our example graph (figures 1 and 3), the nodes $\{c, d, e\}$ together with the edges connecting them form a clique. Cliques represent what social scientists call *primary groups*. A nuclear family is a clique in a social network – often childhood friends also form primary groups or cliques.

Strong Triadic Closure Properties:

In a social network, if F_1 and F_2 are two close friends of yours (connected to you by strong ties), it is likely that F_1 and F_2 are acquaintances (connected to each other by weak ties) – or at least your social network should recommend that F_1 and F_2 explore contacting each other. In particular, if F_1 and F_2 have a large subgroup of common friends (which would include you), it is probable that they are acquaintances – the probability increasing with the size of the set of mutual friends. Thus, you should expect to see lots of cliques of size 3, or try to create as many K_3's as possible by recommending connections among individuals with common close friends.

Note that our simple and intuitive analysis of the diameter of the social networks was based on friends being independent and thus lacking triadic closures. But, a more rigorous analysis will show that in a Poisson random graph (with some triadic closures), a similar analysis will still work!

Back to "strong triadic closure properties"

Triadic Closures: Consider an "augmented" undirected graph, $G = (V, E, E')$, in which $E' \subseteq E$, where E are the edges (also called, ties) and E' represents the set of strong ties.

$$\begin{aligned}(u,v) \in E \quad &\Rightarrow \quad u \text{ and } v \text{ are friends} \\ &\qquad\qquad\qquad \text{(either acquaintances or close friends)} \\ (u,v) \in E' \quad &\Rightarrow \quad u \text{ and } v \text{ are close friends}\end{aligned}$$

The *strong triadic closure property* states that: *if* $(u,v) \in E'$ *and* $(u,w) \in E'$, *then* $(v,w) \in E$. A more probabilistic version states that

$$\begin{aligned}& Pr[(v,w) \in E | (u,v) \in E' \wedge (u,w) \in E'] \\ > \quad & Pr[(v,w) \in E].\end{aligned}$$

That is the knowledge that v and w have a common close friend, u, raises the (conditional) probability that v and w are at least acquaintances. Another way of saying this would be to state that

$$\begin{aligned}& Pr[(v,w) \in E \wedge (u,v) \in E' | (u,w) \in E'] \\ > \quad & Pr[(v,w) \in E \wedge (u,v) \in E' | (u,w) \in E \setminus E'] \ \Box\end{aligned}$$

Mark Granovetter, an American sociologist currently at Stanford University, is widely known for his theory on the spread of information in social networks, known as "The Strength of Weak Ties," discussed in a paper, with the same title and published in 1973. This highly influential sociology paper, with over 19,000 citations, proposed that weak ties enable reaching populations and audiences with much higher efficiency than what is achievable or accessible via strong ties. These findings were later published in the monograph *Getting A Job*, an adaptation of his doctoral dissertation at Harvard University's Department of Social Relations. Recall, from our lecture #1, that when Granovetter surveyed 282 professional, technical, and managerial workers in Newton, MA, he found that of those 282 professionals, those who found jobs through personal contacts (N=54), a substantial percentage 55.6% reported seeing their contact occasionally, and 27.8% rarely. This phenomenon can be modeled through the triadic closure properties as follows: Let us consider a relation R

$$\{(u,v) \in R\} = \text{Event } u \text{ obtained a job through a referral by } v,$$

and

$$Pr[(u,v) \in R] = \text{Probability of } u \text{ obtaining a job through a referral by } v,$$

There are two situations to consider: in both cases, assume that v and w are close friends, and v provides the referral and w is the potential employer. In the first case, when u and v are close friends, w is likely to be an acquaintance and will use information in addition to what v provides in the referral. In the second case, when u and v are just acquaintances, w is unlikely to know u and will go by v's referral only. The argument is that the additional information w may have will lower the probability of u being offered a job.

$$\begin{aligned} & Pr[(u,v) \in R \mid (u,v) \in E'] \\ & \quad = \quad Pr[(u,w) \notin E \wedge (v,w) \in E' \mid (u,v) \in E'] \\ & \quad < \quad Pr[(u,w) \notin E \wedge (v,w) \in E' \mid (u,v) \in E \setminus E'] \\ & \quad = \quad Pr[(u,v) \in R \mid (u,v) \in E \setminus E']. \end{aligned}$$

Next, we will explore how individuals get connected through others in a social network. In a graph, not every pair of vertices is adjacent, but it may be possible to construct "short" sequences of "connecting" vertices from one vertex to another. In that case, the graph will be connected; if it is not, it can be described in terms of a set of connected components.

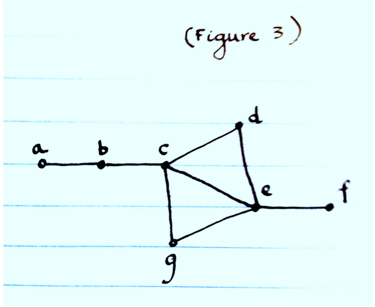


Figure 3: Walks, Trails & Paths.

Walk: A sequence of adjacent vertices $v_0, v_1, \ldots, v_n$ is known as a *walk*. In our example graph (figure 3), the sequence a, b, c, b,

a, b, c is a *walk*. A walk can also be described as a sequence of incident edges, where two edges are said to be incident if they share exactly one vertex.

A walk in which no vertex occurs more than once is known as a *(simple) path*. In our example, the sequence a, b, c, d, e, f is a path. A walk in which no edge occurs more than once is known as a *trail*. In our example, the sequence a, b, c, e, d, c, g is a trail but not a path.

Every path is a trail, and every trail is a walk. A walk is closed if $v_0 = v_n$. A *cycle* can be defined as a closed path in which $n \geq 3$. The sequence c, g, e, d in the example is a cycle.

Connected Component: A *connected component* of a graph is defined as a maximal subgraph in which a *path* exists from every node to every other (i.e., they are mutually reachable). The *size of a component* is defined as the number of nodes it contains. A connected graph has only one connected component.

Tree: A tree is a *connected graph* that contains no cycles. In a tree, every pair of points is connected by a unique path. That is, there is only one way to get from u to v.

The *length of a walk* (and therefore a path or trail) is defined as the number of edges it contains. For example, in our graph, the path a, b, c, d, e has length 4. A walk between two vertices whose length is as short as any other walk connecting the same pair of vertices is called a *geodesic* (the shortest walk/path). Of course, all geodesics are paths. Geodesics are not necessarily unique. From vertex a to vertex f, there are two geodesics: a, b, c, d, e, f and a, b, c, g, e, f.

The graph-theoretic distance (usually shortened to just "distance") between two vertices is defined as the length of a geodesic that connects them. If we compute the distance between every pair of vertices, we can construct a distance matrix D. The maximum distance in a graph defines the graph's *diameter*, Δ. If the graph is not connected, then there exist pairs of vertices that are not mutually reachable so that the distance between them is not defined and the diameter of such a graph is also not defined.

§§§

The powers of a graph's adjacency matrix, A^p (over the integers), give the number of walks of length p between all pairs of nodes. For example, A^2, obtained by multiplying the matrix by itself, has entries that give the number of walks of length 2 that join node v_i to node v_j. Hence, the geodesic distance matrix D has entries $d_{ij} = p$, where p is the smallest p such that $A^p|_{ij} > 0$. (However, there exist much faster algorithms for computing the distance matrix.)

The *eccentricity* $e(v)$ of a point v in a connected graph $G(V, E)$ is $\max d(u, v)$, for all $u \in V$. In other words, a point's eccentricity is equal to the distance from itself to the point farthest away. The eccentricity of node b in our example (figure 3) is 3. The minimum eccentricity of all points in a graph is called the *radius* $\rho(G)$ of the graph, while the maximum eccentricity is the *diameter* $\Delta(G)$ of the graph. In the graph of figure 3, the radius is 2 and the diameter is 4. A vertex that is least distant from all other vertices (in the sense that its eccentricity equals the radius of the graph) is a member of the center of the graph and is called a *central point*. Every tree has a center consisting of either one point or two adjacent points.

Breadth-First Search: For a small graph, we can generally figure out the distance between two nodes by visual inspection; but for larger graphs we need a well-formalized algorithm. The following simple algorithm works well – takes linear amount of work in the number of vertices and edges. Its time complexity is $O(n + m)$.

(0) You are at distance 0 from yourself.

(1) You first declare all of your actual friends to be at distance 1.

(2) You then find all of their friends (not counting people who are already friends of yours), and declare these to be at distance 2.

(3) Then you find all of their friends (again, not counting people who you've already found at distances 1 and 2) and declare these to be at distance 3.

(. . .) Continuing in this way, you search in successive layers, each representing the next distance out. Each new layer is built from all those nodes that (i) have not already been discovered in earlier layers, and that (ii) have an edge to some node in the previous layer.

```
Procedure BFS(G, v)
  Create a queue Q
  ENQUEUE(v, Q)
  mark v; D(v) := 0;
  while Q is non-empty
      t := DEQUEUE(Q)
      for all edges (t, w) in E do
      if w is not marked
          mark w; D(w) := D(t) + 1
          ENQUEUE(w, Q)
```

Figure 4: BFS: Breadth-First Search

This technique is called *breadth-first search*, since it searches the graph outward from a starting node, reaching the closest nodes first. In addition to providing a method of determining distances, it can also serve as a useful conceptual framework to organize the structure of a graph, arranging the nodes based on their distances from a fixed starting point.

Of course, despite the social-network metaphor, the process can be applied to any graph: one just keeps discovering nodes layer-by-layer, building each new layer from the nodes that are connected to at least one node in the previous layer.

A node whose removal from a graph disconnects the graph (or, more generally, increases the number of components in the graph) is called a *cutpoint* or an *articulation point*. The example graph (figure 3) has three cutpoints, namely b, c, and e. A connected, non-trivial graph is called *non-separable* if it has no cutpoints. A *block* or *bi-component* is a maximal nonseparable subgraph. Blocks partition the edges in a graph into mutually exclusive edges. They also share no nodes except cutpoints. Thus, cutpoints decompose graphs into (nearly) non-overlapping sections. In blocks of more than two points, every pair of points lies along a common cycle, which means that there is always a minimum of two ways to get from any point to any other. Our example (figure 3) has the following blocks: $\{a, b\}$, $\{b, c\}$, $\{c, d, e, g\}$, $\{e, f\}$.

The notion of a cutpoint can be generalized to a *cutset*, which is a set of points whose joint removal increases the number of components in the graph. Of particular interest is a *minimum weight cutset*, which is a cutset that is as small as possible (i.e., no other cutset has fewer members). There can be more than one distinct minimum weight cutset in a graph. The size of a graphÕs minimum weight cutset defines the *vertex connectivity* $\kappa(G)$ of a graph, which is the minimum number of nodes that must be removed to increase the number of components in the graph (or render it trivial). The vertex connectivity of a disconnected graph is 0. The vertex connectivity of a graph containing a cutpoint is no higher than 1. The vertex connectivity of a non-separable graph is at least 2. We can analogously define the vertex connectivity $\kappa(u, v)$ of a pair of points u, v as the number of nodes that must be removed to disconnect that pair. The connectivity of the graph $\kappa(G)$ is just the

$$\min_{u,v \in V} \kappa(u, v).$$

Thus, we can think of the point connectivity of a graph as an indicator of the invulnerability of the graph to threats of disconnection by removal of nodes ("unfriending"). If $\kappa(G)$ is high, or if the average $\kappa(u, v)$ is high for all pairs of nodes, then we know that it is fairly difficult to disconnect the nodes in the graph by removing intermediaries.

The vertex-based notions of cutpoint, cutset, vertex connectivity, etc. have analogous counterparts for edges. A *bridge* or *isthmus* is defined as an edge whose removal would increase the number of components in the graph. Edge connectivity is denoted $\lambda(G)$ and the edge connectivity of a pair of nodes is denoted $\lambda(u, v)$. A disconnected graph has $\lambda(G) = 0$, while a graph with a bridge has $\lambda(G) = 1$. Vertex connectivity and edge connectivity are related to each other and to the minimum degree in a graph by WhitneyÕs inequality:

$$\kappa(G) \leq \lambda(G) \leq \delta(G).$$

Social Network Extensions to Graph Theory

Let us look at how graph theory can help us in understanding social networks. Let us start with the notion of *cohesive subsets*.

Cohesive Subsets

It was mentioned earlier that the notion of a *clique* can be seen as formalizing the notion of a *primary group*. A problem with this, however, is that it is too strict to be practical: real groups will contain several pairs of people who don't have a close relationship. A relaxation and generalization of the clique concept is the *n-clique*. There are also the related ideas of n-clans and n-clubs.

***n*-Clique:** An n-clique L of a graph G is a maximal subgraph of G such that for all pairs of vertices u and v of L the distance between them in G is bounded from above by n:

$$\forall_{u,v \in L} d_G(u,v) \leq n.$$

In other words, an n-clique is a set of nodes in which every node can reach every other in n or fewer steps, and the set is maximal in the sense that no other node in the graph is distance n or less from every other node in the subgraph. A 1-clique is the same as an ordinary clique. The set $\{a,b,c,d,e\}$ in our graph (Figure 5 (i)) is an example of a 2-clique. Note that c and e are connected through f, which, however, is not part of the 2-clique. Thus, the path of length n or less linking a member of the n-clique to another member may pass through an intermediary who is not in the group. In this sense, n-cliques are not as cohesive as they might otherwise appear.

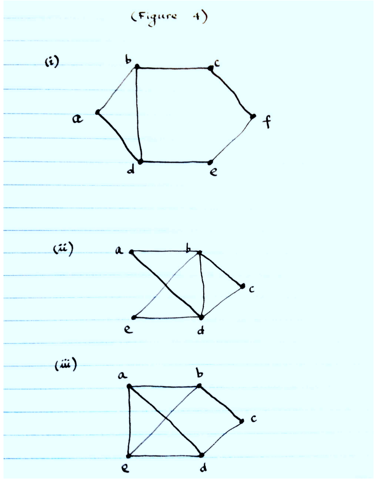


Figure 5: Cliques, Clubs & Clans.

***n*-Club:** An n-club N of a graph G is a maximal subgraph of G of diameter n:

$$\forall_{u,v \in N} d_N(u,v) \leq n.$$

The set $\{a,b,c,d,e\}$ is not a 2-club, since

$$d_G(c,e) = 2, \text{ but } d_L(c,e) = 3.$$

However, $\{a,b,c,d\}$ is a 2-club, but it is not a 2-clique (since it is not the maximal subgraph satisfying 2-clique property).

***n*-Clan:** An n-clan M of a graph G is an n-clique of G such that for all pairs of vertices u and v of M the distance between them in M is bounded from above by n:

$$\forall_{u,v \in M} d_M(u,v) \leq n.$$

In our example, $\{b,c,d,e,f\}$ is a 2-clan. It is a 2-clique with a diameter of 2.

Note that n-clan is both an n-clique and n-club.

Whereas n-cliques, n-clans and n-clubs all generalize the notion of clique via relaxing distance, the *k-plex* generalizes the clique by relaxing density.

***k*-plex** A k-plex is a subset S of nodes such that every member of the set is connected to $n-k$ others, where n is the size of S.

Although not part of the official definition, it is conventional to additionally impose a maximality condition, so that proper subsets of k-plexes are ignored. There are some guarantees on the cohesiveness of k-plexes. For example, k-plexes in which $k < (n+2)/2$ have no distances greater than 2 and cannot contain bridges (making them resistant to attack by deleting an edge). In Figure 5 (i), the set $\{a,b,c,d\}$ fails to be a 2-plex because each member must have at least $4-2=2$ ties to other members of the set, but c has only one tie within the group. In the graph in Figure 5 (ii) and (iii), the set $\{a,b,d,e\}$ are both 2-plexes.

More cohesive than k-plexes are *LS sets*.

LS sets: Let H be a set of nodes in graph $G(V,E)$ and let K be a proper subset of H. Let $\alpha(K)$ denote the number of edges linking members of K to $V-K$ (the set of nodes not in K). Then H is an *LS* set of G if for every proper subset K of H, $\alpha(K) > \alpha(H)$.

The basic idea is that individuals in H have more ties with other members than they do to outsiders.

In the figure 5 (ii), the set $\{a,b,d,e\}$ is not an LS set since $\alpha(\{b,d,e\},\{a\})$ is not greater than $\alpha(\{b,d,e\},\{c\})$. In contrast, the set $\{a,b,d,e\}$ in the figure 5 (iii) does qualify as an LS set.

A key property of LS sets is high edge connectivity.