Social and Information Networks

University of Toronto CSC303
Winter/Spring 2019

Week 4: January 28,30 (2019)

This weeks agenda

Announcements

- Lectures this week by Tyrone Strngway
- Assignment 1 is (or will soon be complete). It is due February 15.
- Last weeks lectures:
- We will complete Chapter 4 of the text with the discussion of Schelling's segregation model (section 4.5).
- The rest of the week will be devoted to Chapter 5 where now negative effects in social networks are introduced.

Is some degree of segregation “natural”?

Chapter 4 ends with a discussion of Schelling’s model that provides an explanation as to how racial neighbourhood segregation can evolve when driven by individuals wanting to be near “people similar to themselves”. Schelling’s model and his simulations lead him to a fundamental observation:

Segregation can and will happen even if there is no explcit individual desire to avoid (say) people of a different race. All that is needed is some desire to be near enough similar people.

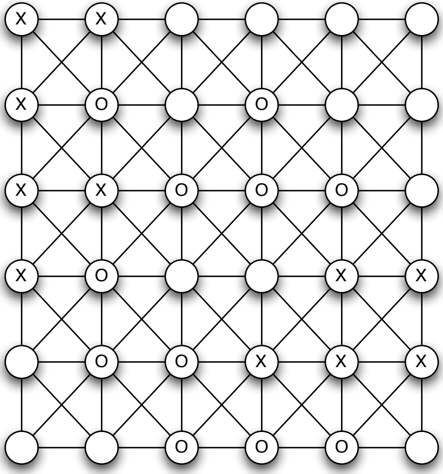
This observation isn’t restricted to racial segregation but we can also witness neghbourhoods that are largely or significantly basd on ethnicity.

In addition to the importance of this fundamental observation, the model provides an interesting study of network dynamics, homophily driven by selection, and how local decisions lead to global structure in a network.

Of course, Schelling’s model does not preclude the presense of other economic and political factors, not does it preclude explicit racism.

The Schelling model

The model itself is quite simple but still hard to analyze analytically. In this model, we view two classes of individuals (X and O) living in a grid. More speficially, individuals occupy some subset of the nodes as depicted in figure 4.15 of the text.



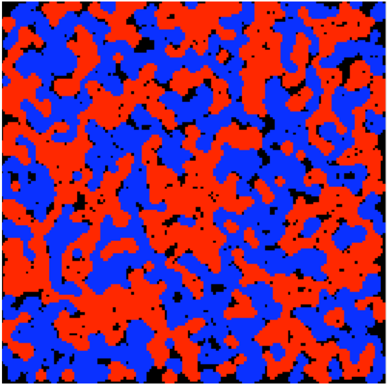
The dynamics of the Schelling model

Schelling then hypotesizes that every individual wants to have at least some threshold t of his/her neighbours to be immediate neighbours. When a individual's thrshold is not met, they move. There are diffeerent versions of the model depending on the order in which individuals move and where they rndomly move to in order to satsify the desire for similarity. The claim is that the results do not change qualitatively.

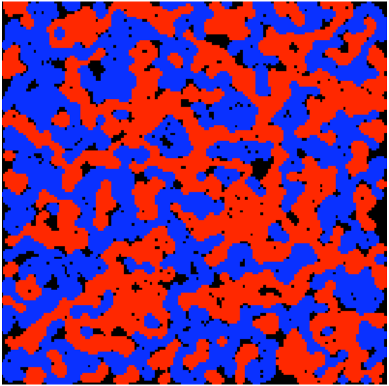
To observe the dynamics, simulatiuons of the network are conducted for different threshold values. What is very apparent is the segegrated structure of the network as it evolves.

The specific gird is a 150 by 15o grid (i.e., 12,500 cells, with 10,000 cells occupied) with both groups equlally represented. The following show the results for thresholds $t = 3$ (i.e. an individual desires less than a majority of his/her neighbours to be similar) and $t = 4$.

Simulations for $t = 3$



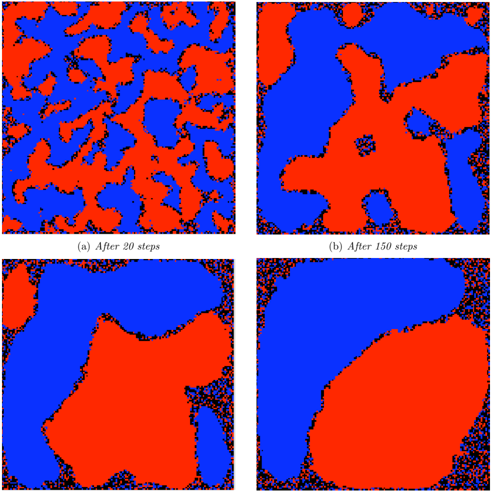
(a) *A simulation with threshold 3.*



(b) *Another simulation with threshold 3.*

Figure 4.17: Two runs of a simulation of the Schelling model with a threshold t of 3, on a 150-by-150 grid with $10,000$ agents of each type. Each cell of the grid is colored red if it is occupied by an agent of the first type, blue if it is occupied by an agent of the second type, and black if it is empty (not occupied by any agent).

Simulation for $t = 4$



(a) *After 20 steps*

(b) *After 150 steps*

(c) *After 350 steps*

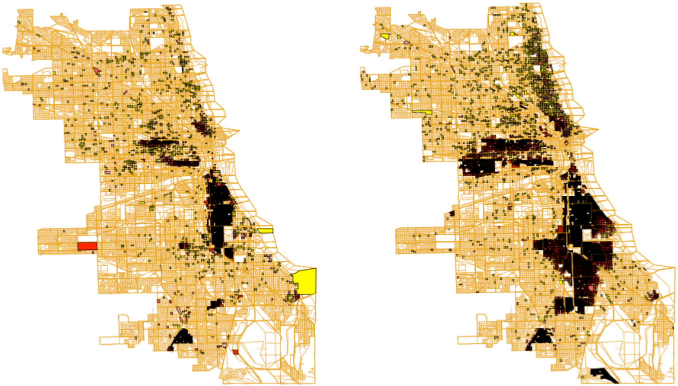
(d) *After 800 steps*

Figure 4.19: Four intermediate points in a simulation of the Schelling model with a threshold t of 4, on a 150-by-150 grid with $10,000$ agents of each type. As the rounds of movement progress, large homogeneous regions on the grid grow at the expense of smaller, narrower regions.

Some concluding comments on the Schelling study

- The model is not constructed so as to build in segregation. More specifically, the model allows for stable configurations thta are well integrated.
- However, given a random starting configuration, the simulations show that people will gravittate to a segregated structure.
- There is a compunding effect of the model dynamics. Namely, when one person leaves, it can result in other neighbours falling below their threshold ansd hence a new desire to leave the current location has been created.
- The word segregation is a term with a very negative connotation due to the use of the term with respect to racial 9(e.g., Jim Crow legislattion) and religious segregation (e.g., ghettos) which was forced by governments. Do we think that neighbourhoods that are concentrated along say ethnic lines is a bad thing? At some level (i.e., Metro Toronto), Toronto may be the most ethically diverse city as is claimed. But at a more detailed level, many neighbours are far from being "integrated".

The reality of neighbourood segregation in Chkicago (1970s)



(a) *Chicago, 1940* (b) *Chicago, 1960*

Stuctural balance when there are positive and negative links

As previously mentioned so far we have restricted attention to social networks where all edges reflect some positive degree of friendship, collaboration, communictation, etc.

Chapter 5 now explores some interesting aspects of networks where edges can be both positive and negatve. This is, of course, quite natural in that people (countries) often have enemies as well as friends (allies). We also have companies that can be aligned in some way or can be competitors. Here the meaning of a edge can seemingly change any general observations more than when there are only positive edges.

Following the development stemming from the distinction between strong and weak ties, we would like to see what we can infer about a network given that some edges are positive and some are negative. More specifically, what can be assumed from certain types of triadic closures? How can local properties (e.g., how edges of a triangle are labelled) can have global implications (i.e., provable results about network structure)?

Some initial assumptions

We start with a strong assumption:
Assume the network is a complete (undirected) graph. That is, as individuals we either like or dislike someone. Furthermore, this is not nuanced in the sense that there is no differentiation as to the of attraction/repulsion).

Later in the chapter, the text considers the issue of networks that are not complete networks. The text also reflects a little on the nature of directed networks (when discussing the *weak balance property*) but essentially this chapter is about undirected networks.

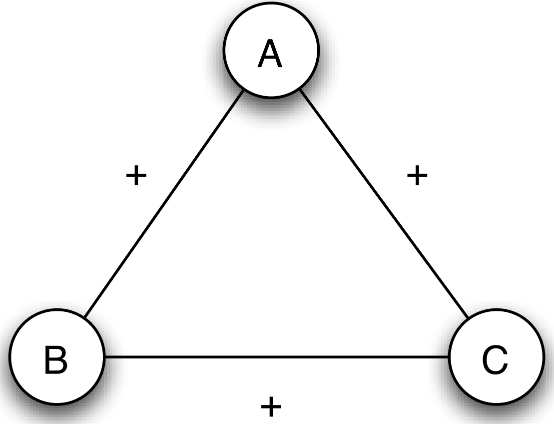
Note: We can assume the graph is connected since otherwise we can consider each connected component separately.

Types of instability

Thinking of networks as people with likes and dislikes of other people (rather than some other possible interpretations), we can consider the 4 different types of labelled triangles in the graph, depending on the number of positive (+) and negative (-) edges. That is, any completely labelled triangle can have 0,1,2, or 3 positive edges and due to the symmetry of a triangle that is all the information we have about any particular triangle.

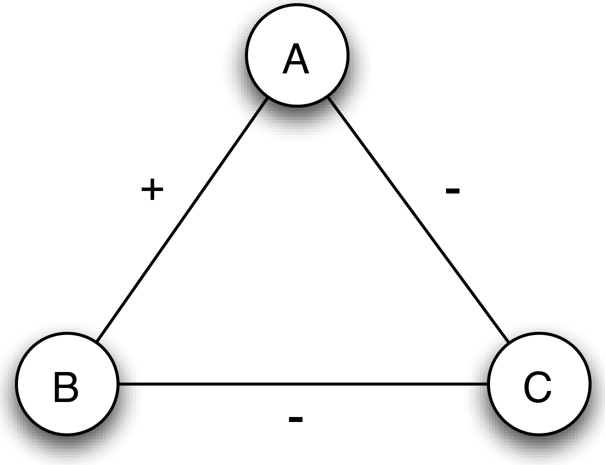
Using a central idea from social psychology, two of the four triangle labellings are considered relatively stable (called *balanced*) and the other two relatively unstable (*not balanced*).
Here follows the four types of triangles as depicted in Figure 5.1 of the text:

A natural stable configuration



In this case, A, B, C are mutual friends and that naturally indicates that they would likely remain so.

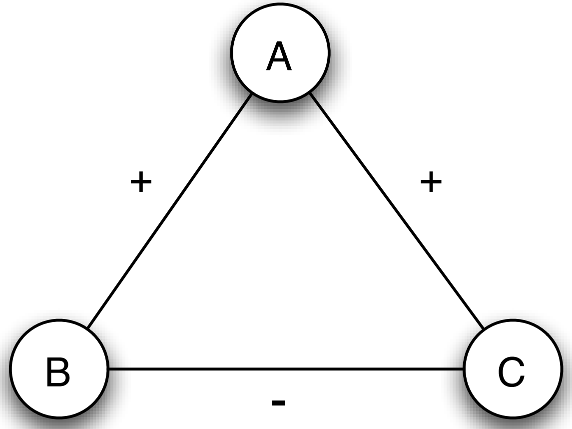
The second stable configuration



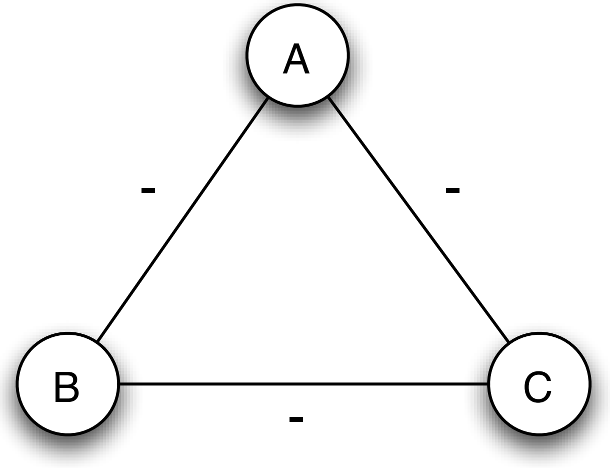
This may be a slightly less obvious stable situation where A and B are friendis and if anything that friendship is reinforced by a mutual dislike for C.

A natural unstable configuration

In this case, A has two friends B and C who unfortunately do not like each other. The claim here is that the stress of this situation will encourage A to either try to have B and C become friends or else for A to take sides with B or C and thus eliminate a friendship so as to move toward the previous stable configuration.

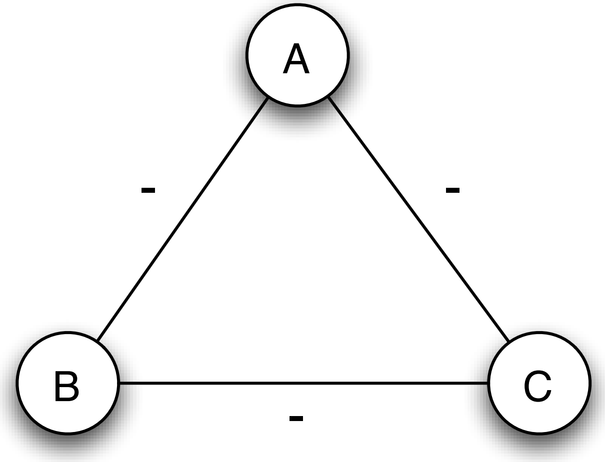


A somewhat less obvious unstable configuration



Why is this unstable

A somewhat less obvious unstable configuration



Why is this unstable The instability here is explained by the phenomena that “the enemy of my enemy becomes my friend” as we sometimes see in international relations.

The strong structural balance property

The underlying behavioural theory is that these unstable triangles cause stress and hence the claim that such unbalnced troangles are not common.

In order to try to ‘understand if this theory tells us anyting about the global structure of the network, we can make the following strong balance assumption (much as we made the strong triadic closure assumption).

Strong structural balance property: Every triangle in the netowrk is balanced.

Recall that we started off with the assumption that the network is a complete graph with every edge labelled so we are assuming a property for all n choose 3 triangles. Of course, we cannt expect this property to hold but just as the strong troadic closure property was an extreme assumption, we can hope that this strong assumotion will also suggest or predict useful information about the network.

Consequence of the strong structural balance property: A provable characterization of networks that satisfy the property

One simple (idealistic) way to construct a network satisfying the property is to asssume that that there are no enemies; everyone is a friend. Is this the only way?

Consequence of the strong structural balance property: A provable characterization of networks that satisfy the property

One simple (idealistic) way to construct a network satisfying the property is to asssume that that there are no enemies; everyone is a friend. Is this the only way?

Suppose that we had two communities of active political people (e.g. $X =$ the "base" for candidate or political party R, , and Y and the "base" for candidate or political party B. In the world of highly politicized politics, it isn't too far of a stretch to think that eveyone within a community are friends and everyone dislikes people in the other community. This kind of network would also clearly satisfy the property.

So far then, we have two possibilities, the network is a clique, or the network is composed of two cliques with a complete bipartite graph of negative edges between the communities.
Are there other possible ways to have the strong balsnce property?

Consequence of the strong structural balance property: A provable characterization of networks that satisfy the property

One simple (idealistic) way to construct a network satisfying the property is to asssume that that there are no enemies; everyone is a friend. Is this the only way?

Suppose that we had two communities of active political people (e.g. $X =$ the "base" for candidate or political party R, , and Y and the "base" for candidate or political party B. In the world of highly politicized politics, it isn't too far of a stretch to think that eveyone within a community are friends and everyone dislikes people in the other community. This kind of network would also clearly satisfy the property.

So far then, we have two possibilities, the network is a clique, or the network is composed of two cliques with a complete bipartite graph of negative edges between the communities.
Are there other possible ways to have the strong balsnce property?

Harary's Balance Theorem

Are there other possible ways to have the strong balance property?

Harary's Balance Theorem

Are there other possible ways to have the strong balance property?

Harary's Balance Theorem

Are there other possible ways to have the strong balance property?

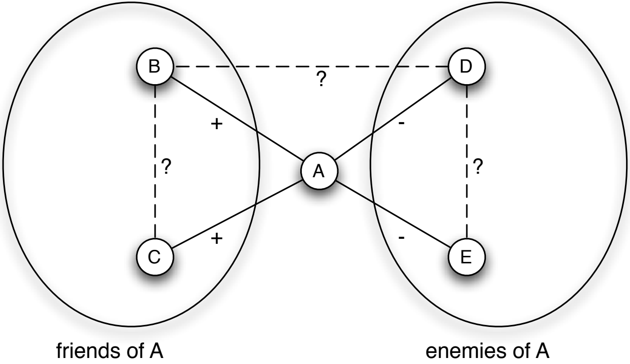
Perhaps suprisingly, thsee two types of networks (no enemies and two opposing communities) are the only possibilities.

This is a theorem and the proof is not difficult as we will show using the figure 5.4 in the text.

Proof
We assume that the network satisfies the strong balance property. If there are no enemies, then we are done. So suppose there is at least one negative edge and for definiteness lets say that edge is adjacent to node A. Let X be all the friends of A and Y all of its enemies. So every node is in either X or Y since every edge is labelled.

Proof of balance theorem continued

Consider the three possible triangles as in the figure. It is easy to see that in order to maintain structural balance, B and C must be friends as must D and E, whereas B abd D (also C and E) must be enemies.



Strong structural balance in networks that are not complete

We will depart form the order of topics in chapter 5, and consider the issue of networks that are not complete. Is there a meaningful sense in which a network is or is not structurally balanced?

One possibility is to ask whether or not there is a way to complete the graph so that it becomes structurally balanced. Of course, if there is already an unbalanced triangle then there is no way to complete the graph into one satisfying the strong balance property.

Aside: Of course, this immediately raises the question as to how many existing edge labels need to be changed so that a complete network is balanced (or an incomplete network can be made to be balanced)? And will networks tend to dynamically evolve into balanced networks. But for now we will assume that all exiting lables are permanent.

How to label missing edges?

Note that when considering the strong triadic property, if all existing triangles satisfied the strong triadic property, then there was always a trivial way to assign labels to unlabelled edges by simply making each unlabelled edge a weak link.

Question: If all existing triangles are balanced, is there a always a way to complete a network so as to form a strongly balanced network?

How to label missing edges?

Note that when considering the strong triadic property, if all existing triangles satisfied the strong triadic property, then there was always a trivial way to assign labels to unlabelled edges by simply making each unlabelled edge a weak link.

Question: If all existing triangles are balanced, is there a always a way to complete a network so as to form a strongly balanced network?

It is easy to see that this is not always possible. For example, comsider a network which is a 4 node cycle having 3 positive edges and one negative edge. Any way to label a "diagonal edge" will lead to an imbalance.

We are then led to the following
Question: Can we determine when there is an efficient algorithm to complete the network so as to satisfy the strong blance property? And if there is a completion, how efficiently can one be found?

Determining when and how to complete a network to satisfy the strong balance property

Clearly, if the existing edges are all poisitve links then there is a trivial way to complete the graph by simply making all missing edges to be positive edges.

So the intersting case is when there are existing negative edges. In this case, the characterization of strongly balanced networks tells us that when the graph is completed, the graph structure must be that of two opposing communities, with only positive edges within each community and only negative edeges for links between the communities.

The previous example of a 4 node cycle is a clue as to how to proceed. That example can be stated as follows: if a network contains a 4 node cycle with one negative edge then it cannot be completed (to be strongly balanced) . More generally, if a network contains a cycle (of any length) with one negative edge, it cannot be completed. And even more generally, if a network contains a cycle having an odd nunber of negative edges it cannot be completed. Why?

The algorithm for determining if a partially labelled network can be completed to the strongly balanced

Let]s call a cycle with an odd number of edges an odd cycle. The desired algorithm will either find an odd cycle (certifying that the netwiork cannot be completed) or it will return a bipartiton of the nodes satisfying the Balance Theorem. This then is also determines if a complete network is balanced.

We proceed as follows:

- Suppose $G = V, E)$ is the given connected network and let $G+ = (V, E^+)$ whee $E^+ = \{e \in E$ such that e is a positive link.$\}$
- We consider the connnected components $\mathcal{C} = C_1, \ldots, C_r$ of G^+ and let $T_1, \ldots, T_r$ be spanning trees for these componenets.
- Note that all edges between any C_i, C_j must be labelled as negative edges (or else they would have beem merged into a larger connected comoponent in G^+..

Completing the algorithm

- Otherwise, consider the graph $G^- = \{\mathcal{C}, E^-\}$ whose nodes are the components of G^+ and whose edges are negative edges in G.
- Since G is connected, G^- is connected.
- if G^- has a cycle with an odd number of negative edges, then by following positive edges in each C_i we have such a cycle in G. We then again have a witness that G cannot be completed.
- Otherwise we are showing that G^- is bipartite and this gives us the biparttition we need for the balance theorem.
- A graph has an odd cycle iff the graph is not bipartite. Brewadth first search can be used to determine whether or not a graph is bipartite (equivalently has a 2-colouring). Hence this development is efficient.

For the remainder of this lecture, we return to the assumption that our networks are undirected complete graphs.

Friends-enemies vs trust-distrust

There is always an ambiguity in social networks as to how to interpret links. Is a friend as we might traditionally mean a “good friend”, or is it a friend as in Facebook friends which in many cases are just people you know? And as we have seen we also use social network links to mean collaboration or communication rather than friendship.

This is both the power of network modeling (i.e., that results can carry over to different settings) and also the danger of misinterpreting results for one type of setting to apply to another.

In chapter 5, we see another instance of the ambiguity where instead of the friend-enemy relation, one can interpret an edge label as a trust-distrust relation.

To what extent should we expect intuiton for friendship to carry over to trust? As discussed in the text, one distinction between these settings is that turst may be more of a directed edge concept relative to friendship. (Of course, even for friendship the relation may not be symmetric which is why maybe we should reserve the term of “friend” for a good friend.)

The ambiguity in the trust-distrust relation

Ignoring the fact that trust might not be at all symmetric, there is an additional ambiguity in the trust-distrust terminology. Namely, the text considers two possible interpretations that are meaningful even in the context of a single setting as in the online product rating site Epionions.

1. If trust is aligned with agrement on political issues as in the ratings of political commentary, then the four cases of balanced and unbalanced triangles still seem to apply. In particular, if A distrusts B and B distrusts C, it is reasonable to assume that A trusts C and hence a triangle having three negative labels is not stable.
2. However, if A disttrusts B is aligned with A believing that he/she is more knowledgeable than B about a certain product, then a triangle having three negative labels is stable.

This suggests that it is reasonable to study a weaker form of structural balance.

A weaker form of structural balance

It is then interesting to consider a weaker form of structural balance where the only unstable triangles are those having two positive labels.

This leades to the following definition (analogous to the strong structural balance property:
A network satisfies the weak structural balance property if it does not contain any triangles with exactly two positive edges. This in turn leads to the following
Question: Is there a characterization of which (complete) networks satisfy the weak structural balance property?

A weaker form of structural balance

It is then interesting to consider a weaker form of structural balance where the only unstable triangles are those having two positive labels.

This leades to the following definition (analogous to the strong structural balance property:
A network satisfies the weak structural balance property if it does not contain any triangles with exactly two positive edges. This in turn leads to the following
Question: Is there a characterization of which (complete) networks satisfy the weak structural balance property?

Since every network that satisfies the strong balance property must also satisfy the weak balance property, the characterization of strongly balanced networks must be a special case of weakly balanced networks. Indeed we have the following characterization:

Theorem: A network $G = (V, E)$ satisfies the weak strucrtural balance property iff $V = V_1 \cup V_2 \ldots V_r$ such that all edges within any V_i are positive edges and all edges between V_i and V_j $(i \neq j)$ are negative edges.

Proof of the characterization of weak structural balance

Clearly if the network $G = (V, E)$ has the network structure specified in the Theorem, then the network satisfies the weak balance property. The converse (that the weak balancet property implies the network structure) is a reasonably simple inductive argument (say with respect to the number of edges).

Consider any node A and let X be all the friends of A.
The following two claims are easy to verify:

- Any $B, C \in X$ are friends
- If $B \in X$ and $D \notin X$, then B and D are enemies.

Upon removing the nodes in X, the induced network G' of the remaining nodes still must satisfy the weak structure balance property and hence by the induction hypothesis must have the stated network structure.