Social and Information Networks

CSCC46H, Fall 2019
Lecture 4

Prof. Ashton Anderson
ashton@cs.toronto.edu
Logistics

A1 due next Monday @ 12pm on MarkUs

First letter of last name A-H? First blog post due this Friday
Today

Signed networks
Empirical phenomena in networks
Positive and Negative Relationships

So far, edges mostly interpreted positively
—Friendship
—Interaction
—Collaboration

But relationships can be negative too
—Dislike
—Bad interaction
—Enemy
Network Representation

How would you model this?
Signed Networks

Networks with positive and negative relationships

Consider an undirected complete graph
Label each edge as either:

Positive: friendship, trust, positive sentiment, …

Negative: enemy, distrust, negative sentiment, …
Questions about Signed Networks

What are the **typical patterns of interaction** in signed networks?

How do we **reason** about **local and global structure** of positive and negative interactions?

What are the **patterns in empirical data**?
Signed Networks

Networks with positive and negative relationships

Our basic unit of investigation will be signed triangles

Focus on undirected networks
5.1. STRUCTURAL BALANCE

A, B, and C are mutual friends: balanced.

A is friends with B and C, but they don’t get along with each other: not balanced.

A and B are friends with C as a mutual enemy: balanced.

A, B, and C are mutual enemies: not balanced.

Figure 5.1: Structural balance: Each labeled triangle must have 1 or 3 positive edges.
Structural Balance

5.1. STRUCTURAL BALANCE

(a) A, B, and C are mutual friends: balanced.

(b) A is friends with B and C, but they don't get along with each other: not balanced.

(c) A and B are friends with C as a mutual enemy: balanced.

(d) A, B, and C are mutual enemies: not balanced.

Figure 5.1: Structural balance: Each labeled triangle must have 1 or 3 positive edges.

• Similarly, there are sources of instability in a configuration where each of A, B, and C are mutual enemies (as in Figure 5.1(d)). In this case, there would be forces motivating two of the three people to “team up” against the third (turning one of the three edge labels to a +).

Based on this reasoning, we will refer to triangles with one or three +’s as balanced, since they are free of these sources of instability, and we will refer to triangles with zero or two +’s as unbalanced. The argument of structural balance theorists is that because unbalanced triangles are sources of stress or psychological dissonance, people strive to minimize them in their personal relationships, and hence they will be less abundant in real social settings than...
Structural Balance

5.1. STRUCTURAL BALANCE

(a) A, B, and C are mutual friends: balanced.

(b) A is friends with B and C, but they don't get along with each other: not balanced.

(c) A and B are friends with C as a mutual enemy: balanced.

(d) A, B, and C are mutual enemies: not balanced.

Figure 5.1: Structural balance: Each labeled triangle must have 1 or 3 positive edges.

Friends (thus turning the B - C edge label to +); or else for A to side with one of B or C against the other (turning one of the edge labels out of A to a +).

Similarly, there are sources of instability in a configuration where each of A, B, and C are mutual enemies (as in Figure 5.1(d)). In this case, there would be forces motivating two of the three people to "team up" against the third (turning one of the three edge labels to a +).

Based on this reasoning, we will refer to triangles with one or three +'s as balanced, since they are free of these sources of instability, and we will refer to triangles with zero or two +'s as unbalanced. The argument of structural balance theorists is that because unbalanced triangles are sources of stress or psychological dissonance, people strive to minimize them in their personal relationships, and hence they will be less abundant in real social settings than...

Four signed triads: which are stable?
Theory of Structural Balance

Start with the intuition [Heider ’46]:

Friend of my friend is my friend
Enemy of enemy is my friend
Enemy of friend is my enemy

Look at connected triples of nodes:

- Balanced: Consistent with “friend of a friend” or “enemy of the enemy” intuition
- Unbalanced: Inconsistent with the “friend of a friend” or “enemy of the enemy” intuition
Structural Balance

Which network is balanced?

122

CHAPTER 5. POSITIVE AND NEGATIVE RELATIONSHIPS

A

C

D

B

A

C

D

B

Figure 5.2: The labeled four-node complete graph on the left is balanced; the one on the right is not.

Defining Structural Balance for Networks.

So far we have been talking about structural balance for groups of three nodes. But it is easy to create a definition that naturally generalizes this to complete graphs on an arbitrary number of nodes, with edges labeled by '+'s and '-'s.

Specifically, we say that a labeled complete graph is balanced if every one of its triangles is balanced — that is, if it obeys the following:

Structural Balance Property: For every set of three nodes, if we consider the three edges connecting them, either all three of these edges are labeled '+', or else exactly one of them is labeled '+'.

For example, consider the two labeled four-node networks in Figure 5.2. The one on the left is balanced, since we can check that each set of three nodes satisfies the Structural Balance Property above. On the other hand, the one on the right is not balanced, since among the three nodes A, B, C, there are exactly two edges labeled '+', in violation of Structural Balance. (The triangle on B, C, D also violates the condition.)

Our definition of balanced networks here represents the limit of a social system that has eliminated all unbalanced triangles. As such, it is a fairly extreme definition — for example, one could instead propose a definition which only required that at least some large percentage of all triangles were balanced, allowing a few triangles to be unbalanced. But the version with all triangles balanced is a fundamental first step in thinking about this concept; and
Balanced/Unbalanced Networks

**Define:** A complete graph is *balanced* if every connected triple of nodes has:

- All 3 edges labeled +  \quad \text{or}  \quad \text{Exactly 1 edge labeled +}

![Graph Example](image1.png) \quad \text{Unbalanced}  

![Graph Example](image2.png) \quad \text{Balanced}
The Tribes of Eastern Central Highlands of New Guinea
How general is this?

At a general level, what does a balanced network (i.e., a balanced labeled complete graph) look like? Given any specific example, we can check all triangles to make sure that they each obey the balance conditions; but it would be much better to have a simple conceptual description of what a balanced network looks like in general.

One way for a network to be balanced is if everyone likes each other; in this case, all triangles have three + labels. On the other hand, the left-hand side of Figure 5.2 suggests a slightly more complicated way for a network to be balanced: it consists of two groups of friends (A, B and C, D), with negative relations between people in different groups. This is actually true in general: suppose we have a labeled complete graph in which the nodes can be divided into two groups, \( X \) and \( Y \), such that every pair of nodes in \( X \) like each other, every pair of nodes in \( Y \) like each other, and everyone in \( X \) is the enemy of everyone in \( Y \). (See the schematic illustration in Figure 5.3.) You can check that such a network is balanced: a triangle contained entirely in one group or the other has three + labels, and a triangle with two people in one group and one in the other has exactly one + label.

So this describes two basic ways to achieve structural balance: either everyone likes each other; or the world consists of two groups of mutual friends with complete antagonism...
Local Balance $\rightarrow$ Global Factions

The Balance Theorem: Balance implies global coalitions
[Cartwright-Harary]

If all triangles are balanced, then either:

A) The network contains only positive edges, or

B) The network can be split into two factions: Nodes can be split into 2 sets where negative edges only point between the sets.
Balance Theorem

Global coalitions => balance
Straightforward
Every complete graph that looks like “this” is balanced

Balance => Global coalitions
Less straightforward
Every complete graph that’s balanced looks like “this”?
Balance Theorem

Global coalitions => balance:

Any triangle is one of two types:
A) All 3 nodes in one of the partitions
B) 2 nodes in one partition, 1 in the other

A): all 3 edges are +  →  balanced
B): 2 nodes in one partition are +,
    other 2 edges are -  →  balanced
Proof of Balance Theorem

Balance $\Rightarrow$ Global coalitions:

Pick a node $A$.

Because it’s a complete graph, $A$ is either friends or enemies with each person.

Now **check 3 cases:**

![Diagram showing relationships between nodes A, B, C, D, and E.]
Proof of Balance Theorem

Every node in L is enemy of R

Any 2 nodes in L are friends

Any 2 nodes in R are friends
Balance Theorem

Global coalitions $\Rightarrow$ balance

**Straight-forward**

Every complete graph partitioned into two friendly coalitions that dislike either other is balanced

Balance $\Rightarrow$ Global coalitions

**Less straight-forward**

Every complete graph that’s balanced can be partitioned into two friendly coalitions that dislike either other
European alliances, pre-WWI

(a) Three Emperors’ League 1872–81
(b) Triple Alliance 1882
(c) German-Russian Lapse 1890
(d) French-Russian Alliance 1891–94
(e) Entente Cordiale 1904
(f) British Russian Alliance 1907
International relations:

Positive edge: alliance

Negative edge: animosity

Separation of Bangladesh from Pakistan in 1971: **US** supports **Pakistan**. Why?

**USSR** was the enemy of **China**

**China** was the enemy of **India**

**India** was the enemy of **Pakistan**

**US** was friendly with **China**

**China** vetoed **Bangladesh** from U.N.
5.1. STRUCTURAL BALANCE

A, B, and C are mutual friends: balanced.

A is friends with B and C, but they don't get along with each other: not balanced.

A and B are friends with C as a mutual enemy: balanced.

A, B, and C are mutual enemies: not balanced.

Figure 5.1: Structural balance: Each labeled triangle must have 1 or 3 positive edges.

Based on this reasoning, we will refer to triangles with one or three +'s as balanced, since they are free of these sources of instability, and we will refer to triangles with zero or two +'s as unbalanced. The argument of structural balance theorists is that because unbalanced triangles are sources of stress or psychological dissonance, people strive to minimize them in their personal relationships, and hence they will be less abundant in real social settings than What if we allow three mutual enemies?
Weak Structural Balance →
Many Global Factions

**Define:** A complete network is *weakly balanced* if there is no triangle with exactly 2 positive edges and 1 negative edge.

**Characterization** of Weakly Balanced Networks:
If a labeled complete graph is weakly balanced, then its nodes can be **partitioned**
(divided into groups such that two nodes belonging to the same group are friends, and every two nodes belonging to different groups are enemies)

**Global picture:** same thing as before, but with many factions, not necessarily two
Proof of Characterization

Pick a node A.

Because it’s a complete graph, A is either friends or enemies with each person.

Now check 2 cases:
Proof of Characterization

All of A’s friends are friends with each other and are enemies with all of A’s enemies

Remove A and his friends from the graph and recurse!

Graph still weakly balanced, find a second group, same argument applies, recurse until we’ve found all factions
Balance in General Networks

So far we’ve talked about complete graphs.

What about incomplete graphs?
Signed Graph: Is it Balanced?
Balance in General Networks

So far we talked about complete graphs

**Def 1: Local view**

Fill in the missing edges to achieve balance

If the graph is “Balance-able”, then call it balanced
Balance in General Networks

So far we talked about complete graphs

Def 1: Local view
Fill in the missing edges to achieve balance

If the graph is “Balance-able”, then call it balanced
So far we talked about complete graphs.

**Def 2: Global view**

Divide the graph into two coalitions.

If you can separate the graph into coalitions as before, call it balanced.
Balance in General Networks

So far we talked about complete graphs

Def 1: Local view
Fill in the missing edges to achieve balance

Def 2: Global view
Divide the graph into two coalitions

The 2 definitions are equivalent!
Balance in General Networks

Claim: in general (not necessarily complete) networks, the **local** and **global** definitions of balance are equivalent.

**Def 1: Local view**
Fill in the missing edges to achieve balance.

**Def 2: Global view**
Divide the graph into two coalitions.
Balance in General Networks

**Actually easy to see:**

**Local => global:** (if you can fill in edges such that the resulting complete graph is balanced, then it can be divided into coalitions)

After filling in, we have a complete network as before, the Balance Theorem applies
Balance in General Networks

Actually easy to see:

Global => local: (if the graph can be divided into coalitions, then you can fill in edges that results in a complete balanced graph)

Fill in edges within and between coalitions as before: positive edges within the coalitions and negative edges between them
Balance in General Networks

Actually easy to see:

Local $\Rightarrow$ global: after filling in, result in complete network as before

Global $\Rightarrow$ local: fill in edges within and between coalitions as before

Done!
Balance in General Networks

We have a natural definition for balance in general signed networks

“Natural” because we arrived at it two different ways that turn out to be equivalent
Balance in General Networks

We have a natural definition for **balance** in general signed networks

“**Natural**” because we arrived at it **two different ways** that turn out to be equivalent

But, there’s a problem: **how to actually check if a network is balanced in this way?**
Balance in General Networks

Why isn’t this graph balanced?
Balance in General Networks

Why isn’t this graph balanced?

Walk around a cycle, every time we see a negative edge we have to switch coalitions.
Is a Signed Network Balanced?

**Theorem:** Graph is balanced if and only if it contains no cycle with an odd number of negative edges

[Harary 1953, 1956]
Is a Signed Network Balanced?

**Theorem**: Graph is **balanced** if and only if it contains **no cycle with an odd number of negative edges** [Harary 1953, 1956]

This theorem is saying that the **only way** a graph can be unbalanced is if there is **a cycle with an odd number of negative cycles**. That’s the only possible problem!
Is a Signed Network Balanced?

**Theorem:** Graph is **balanced** if and only if it contains no cycle with an odd number of negative edges

[Harary 1953, 1956]

**Proof:** We will show that every graph is either balanced or contains a cycle with odd number of negative edges
Is a Signed Network Balanced?

Theorem: Graph is balanced if and only if it contains no cycle with an odd number of negative edges [Harary 1953, 1956]

Proof by algorithm: We will do this by actually constructing an algorithm that either outputs a division into coalitions or a cycle with odd number of negative edges.

Because these are the only two outcomes, this proves the claim.
Is a Signed Network Balanced?

Theorem: Graph is **balanced** if and only if it contains no cycle with an odd number of negative edges [Harary 1953, 1956]

Proof sketch: Our algorithm will try to assign nodes to coalitions such that the graph is balanced. We will reason that the only way it can fail is if there is a cycle with an odd number of negative edges.
Is a Signed Network Balanced?

Signed graph algorithm:

**Step 1:** Find connected components on + edges and for each component create a super-node
  - Since nodes connected by a + edge must be in same coalition
  - If any – edge in the super node, done (cycle with 1 negative edge)

**Step 2:** Connect components A and B if there is a negative edge between the members
  - Note there are only negative edges pointing out of a super-node (otherwise should’ve connected the two super-nodes that have a positive edge)
Is a Signed Network Balanced?

**Signed graph algorithm**

- Now we have a graph on super-nodes joined by negative edges
- Just need to consistently assign super-nodes to coalitions X and Y
- BFS starting at any node in the super-node graph (which only has – edges)
- Produces a set of layers of increasing distances from the root
- Call all even layers X and odd layers Y
- If edges are only between adjacent layers (not within-layer), then all – edges point between X and Y, **balanced**!
- Otherwise, within-layer edge A-B. Cycle G-A-B-G has length 2k+1, therefore it’s odd, therefore **unbalanced**!
Is a Signed Network Balanced?

Two outcomes:

1) label each super-node as either X or Y, in such a way that every edge has endpoints with opposite labels. Then we can create a balanced division of the original graph, by labeling each node the way its supernode is labeled in the reduced graph.

2) find a cycle in the original graph that has an odd number of negative edges. Simply “stitch together” these negative edges using paths consisting entirely of positive edges that go through the insides of the supernodes.
Signed Graph: Is it Balanced?
Positive Connected Components
Reduced Graph on Super-Nodes
BFS on Reduced Graph

Using BFS assign each node a side

Graph is **unbalanced** if any two connected super-nodes are assigned the same side

![Graph Diagram]

Unbalanced!
Where Do Signed Edges Come From?

In many online applications users express positive and negative attitudes/opinions:

- Through **actions**:  
  - Rating a product/person
  - Pressing a “like” button

- Through **text**:  
  - Writing a comment, a review

- **Success of these online applications is built on people expressing opinions**
  - Recommender systems
  - Wisdom of the Crowds
  - Sharing economy
Global Structure of Signed Nets

Intuitive picture of social network in terms of densely linked clusters

How does structure interact with links?

Embeddedness of link (A,B): Number of shared neighbors
Global Factions: Embeddedness

Embeddedness of ties: Positive ties tend to be more embedded

Graph and data plots showing the fraction of plus edges against the number of common neighbors for Epinions and Wikipedia datasets.
Real Large Signed Networks

Each link A-B is **explicitly** tagged with a sign:

**Epinions:** Trust/Distrust
- Does A trust B’s product reviews?
  (only positive links are visible to users)

**Wikipedia:** Support/Oppose
- Does A support B to become Wikipedia administrator?

**Slashdot:** Friend/Foe
- Does A like B’s comments?

**Other examples:**
- Online multiplayer games

<table>
<thead>
<tr>
<th></th>
<th>Epinions</th>
<th>Slashdot</th>
<th>Wikipedia</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nodes</td>
<td>119,217</td>
<td>82,144</td>
<td>7,118</td>
</tr>
<tr>
<td>Edges</td>
<td>841,200</td>
<td>549,202</td>
<td>103,747</td>
</tr>
<tr>
<td>+ edges</td>
<td>85.0%</td>
<td>77.4%</td>
<td>78.7%</td>
</tr>
<tr>
<td>- edges</td>
<td>15.0%</td>
<td>22.6%</td>
<td>21.2%</td>
</tr>
</tbody>
</table>
### Balance in Our Network Data

**Does structural balance hold?**

Compare frequencies of signed triads in real and “shuffled” signs

<table>
<thead>
<tr>
<th>Triad</th>
<th>Epinions</th>
<th>Wikipedia</th>
<th>Consistent with Balance?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$P(T)$</td>
<td>$P_0(T)$</td>
<td>$P(T)$</td>
</tr>
<tr>
<td>Balanced</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+ + +</td>
<td>0.87</td>
<td>0.62</td>
<td>0.70</td>
</tr>
<tr>
<td>- - -</td>
<td>0.07</td>
<td>0.05</td>
<td>0.21</td>
</tr>
<tr>
<td>Unbalanced</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+ + -</td>
<td>0.05</td>
<td>0.32</td>
<td>0.08</td>
</tr>
<tr>
<td>- - -</td>
<td>0.007</td>
<td>0.003</td>
<td>0.011</td>
</tr>
</tbody>
</table>

$P(T)$ … fraction of a triad

$P_0(T)$ … triad fraction if the signs would appear at random
Homophily

“Birds of a Feather Flock Together”
Homophily

4.1. HOMOPHILY

Figure 4.1: Homophily can produce a division of a social network into densely-connected, homogeneous parts that are weakly connected to each other. In this social network from a town’s middle school and high school, two such divisions in the network are apparent: one based on race (with students of different races drawn as differently colored circles), and the other based on friendships in the middle and high schools respectively [304].

B and C have a common friend A, then there are increased opportunities and sources of trust on which to base their interactions, and A will also have incentives to facilitate their friendship. However, social contexts also provide natural bases for triadic closure: since we know that A-B and A-C friendships already exist, the principle of homophily suggests that B and C are each likely to be similar to A in a number of dimensions, and hence quite possibly similar to each other as well. As a result, based purely on this similarity, there is an elevated chance that a B-C friendship will form; and this is true even if neither of them is aware that the other one knows A.

The point isn’t that any one basis for triadic closure is the “correct” one. Rather, as we take into account more and more of the factors that drive the formation of links in a social network, we find that homophily is a powerful force in these developments.

• US middle school + high school
• node color = self-identified race
Homophily: Age

- Facebook friendship network, 2011
Homophily: Nationality

- Facebook friendship network, 2011

Figure 9. Normalized country adjacency matrix. Matrix of edges between countries with >1 million users and >50% Facebook penetration shown on a log scale. To normalize, we divided each element of the adjacency matrix by the product of the row country degree and column country degree. The data shows that 84.2% of edges are within countries. So the network divides fairly cleanly along country lines into network clusters or communities. This mesoscopic-scale organization is to be expected as Facebook captures social relationships divided by national borders. We can further quantify this division using the modularity $Q$ which is the fraction of edges within communities minus the expected fraction of edges within communities in a random version of the network that preserves the degrees for each individual but is otherwise random. In this case, the communities are the countries. The computed value is $Q = 0.7486$ which is quite large and indicates a strongly modular network structure at the scale of countries. Especially considering that unlike numerous studies using the modularity to detect communities, we in no way attempted to maximize it directly, and instead merely utilized the given countries as community labels. We visualize this highly modular structure in Fig. 9. The figure displays a heatmap of the number of edges between the 54 countries where the active Facebook user population exceeds one million users and is more than 50% of the internet-enabled population. To be entirely accurate, the shown matrix contains each edge twice, once in both directions, and therefore has twice the number of edges in diagonal elements. The number of edges was normalized by dividing the $ij$th entry by the row and column sums, equal to the product of the degrees of country $i$ and $j$. The ordering of the countries was then determined via complete linkage hierarchical clustering.
Homophily: Friend count

- Facebook friendship network, 2011
Homophily

• Connections don’t form uniformly at random

• **Null model**: what if they were forming at random?

• **Measuring homophily**: are there fewer connections between nodes across traits than you’d expect at random?

• **Homophily test**: If the fraction of cross-gender edges is significantly less than at random, then there is evidence of homophily.
CHAPTER 4. NETWORKS IN THEIR SURROUNDING CONTEXTS

Figure 4.2: Using a numerical measure, one can determine whether small networks such as this one (with nodes divided into two types) exhibit homophily.

And ultimately, one expects most links to in fact arise from a combination of several factors—partly due to the effect of other nodes in the network, and partly due to the surrounding contexts.

Measuring Homophily.

When we see striking divisions within a network like the one in Figure 4.1, it is important to ask whether they are “genuinely” present in the network itself, and not simply an artifact of how it is drawn. To make this question concrete, we need to formulate it more precisely: given a particular characteristic of interest (like race, or age), is there a simple test we can apply to a network in order to estimate whether it exhibits homophily according to this characteristic?

Since the example in Figure 4.1 is too large to inspect by hand, let’s consider this question on a smaller example where we can develop some intuition. Let’s suppose in particular that we have the friendship network of an elementary-school classroom, and we suspect that it exhibits homophily by gender: boys tend to be friends with boys, and girls tend to be friends with girls. For example, the graph in Figure 4.2 shows the friendship network of a (small) hypothetical classroom in which the three shaded nodes are girls and the six unshaded nodes are boys. If there were no cross-gender edges at all, then the question of homophily would be easy to resolve: it would be present in an extreme sense. But we expect that homophily should be a more subtle effect that is visible mainly in aggregate—as it is, for example, in the real data from Figure 4.1. Is the picture in Figure 4.2 consistent with homophily?

There is a natural numerical measure of homophily that we can use to address questions:

\[ p = \text{Probability that a node is white} \]
\[ q = \text{Probability that a node is red} \]

Prob an edge is between two white nodes?
Prob an edge is between two red nodes?
Prob an edge is between 1 red, 1 white?

Homophily test:
Homophily

\[ p = \text{Probability that a node is white} \quad 6/9 = 2/3 \]
\[ q = \text{Probability that a node is red} \quad 3/9 = 1/3 \]

Prob an edge is between two white nodes? \[ p^2 \]
Prob an edge is between two red nodes? \[ q^2 \]
Prob an edge is between 1 red, 1 white? \[ 2pq \]

Homophily test: \[ 2pq = 4/9 = 8/18 \]

Observed: 5/18