Political Network Survey Data

Betsy Sinclair (WUSTL)

October 2018

Substitute Teacher

- I am the substitute teacher today. Jacob Montgomery is attending to a family emergency.
- Here are the ground rules for today:
    1. Laptops fine but pay attention because otherwise I'll wish Jacob had simpy cancelled class.
    2. Lots of opportunities for engagement.
    3. Happy to be interrupted.

Substitute Teacher

- I am the substitute teacher today. Jacob Montgomery is attending to a family emergency.
- Here are the ground rules for today:
    1. Laptops fine but pay attention because otherwise I'll wish Jacob had simpy cancelled class.
    2. Lots of opportunities for engagement.
    3. Happy to be interrupted.

Substitute Teacher

- I am the substitute teacher today. Jacob Montgomery is attending to a family emergency.
- Here are the ground rules for today:
  1. Laptops fine but pay attention because otherwise I'll wish Jacob had simpy cancelled class.
  2. Lots of opportunities for engagement.
  3. Happy to be interrupted.

Substitute Teacher

- I am the substitute teacher today. Jacob Montgomery is attending to a family emergency.
- Here are the ground rules for today:
  1. Laptops fine but pay attention because otherwise I'll wish Jacob had simpy cancelled class.
  2. Lots of opportunities for engagement.
  3. Happy to be interrupted.

Substitute Teacher

- I am the substitute teacher today. Jacob Montgomery is attending to a family emergency.
- Here are the ground rules for today:
    1. Laptops fine but pay attention because otherwise I'll wish Jacob had simpy cancelled class.
    2. Lots of opportunities for engagement.
    3. Happy to be interrupted.

Biography

- Earned my phD from Caltech in 2007 in "social science".
- Mike Alvarez, Jonathan Katz, and Jonathan Nagler formed my political science committee
- First job at the University of Chicago, moved to WUSTL with tenure in 2012. Full professor in 2018.
- Married to fellow Caltech grad and microeconomic theorist Brian Rogers. Three kids (James, age 6 and Miles and Charlie, age 3).
- Research in American political behavior and political methodology
- CV

Biography

- Earned my phD from Caltech in 2007 in "social science".
- Mike Alvarez, Jonathan Katz, and Jonathan Nagler formed my political science committee
- First job at the University of Chicago, moved to WUSTL with tenure in 2012. Full professor in 2018.
- Married to fellow Caltech grad and microeconomic theorist Brian Rogers. Three kids (James, age 6 and Miles and Charlie, age 3).
- Research in American political behavior and political methodology
- CV

Biography

- Earned my phD from Caltech in 2007 in "social science".
- Mike Alvarez, Jonathan Katz, and Jonathan Nagler formed my political science committee
- First job at the University of Chicago, moved to WUSTL with tenure in 2012. Full professor in 2018.
- Married to fellow Caltech grad and microeconomic theorist Brian Rogers. Three kids (James, age 6 and Miles and Charlie, age 3).
- Research in American political behavior and political methodology
- CV

Biography

- Earned my phD from Caltech in 2007 in "social science".
- Mike Alvarez, Jonathan Katz, and Jonathan Nagler formed my political science committee
- First job at the University of Chicago, moved to WUSTL with tenure in 2012. Full professor in 2018.
- Married to fellow Caltech grad and microeconomic theorist Brian Rogers. Three kids (James, age 6 and Miles and Charlie, age 3).
- Research in American political behavior and political methodology
- CV

Biography

- Earned my phD from Caltech in 2007 in "social science".
- Mike Alvarez, Jonathan Katz, and Jonathan Nagler formed my political science committee
- First job at the University of Chicago, moved to WUSTL with tenure in 2012. Full professor in 2018.
- Married to fellow Caltech grad and microeconomic theorist Brian Rogers. Three kids (James, age 6 and Miles and Charlie, age 3).
- Research in American political behavior and political methodology
- CV

Biography

- Earned my phD from Caltech in 2007 in "social science".
- Mike Alvarez, Jonathan Katz, and Jonathan Nagler formed my political science committee
- First job at the University of Chicago, moved to WUSTL with tenure in 2012. Full professor in 2018.
- Married to fellow Caltech grad and microeconomic theorist Brian Rogers. Three kids (James, age 6 and Miles and Charlie, age 3).
- Research in American political behavior and political methodology
- CV

Goals for today

- Discuss the frontier of the networks field with an eye towards what can be gleaned from survey research.
- Focus on the current set of best practices with respect to measurement and inferential strategies that are possible in terms of estimating network effects.
- Present open questions and talk about what is exciting.
- For fun: Interdisciplinary coauthorships place projects in better journals

Goals for today

- Discuss the frontier of the networks field with an eye towards what can be gleaned from survey research.
- Focus on the current set of best practices with respect to measurement and inferential strategies that are possible in terms of estimating network effects.
- Present open questions and talk about what is exciting.
- For fun: Interdisciplinary coauthorships place projects in better journals

Goals for today

- Discuss the frontier of the networks field with an eye towards what can be gleaned from survey research.
- Focus on the current set of best practices with respect to measurement and inferential strategies that are possible in terms of estimating network effects.
- Present open questions and talk about what is exciting.
- For fun: Interdisciplinary coauthorships place projects in better journals

Goals for today

- Discuss the frontier of the networks field with an eye towards what can be gleaned from survey research.
- Focus on the current set of best practices with respect to measurement and inferential strategies that are possible in terms of estimating network effects.
- Present open questions and talk about what is exciting.
- For fun: Interdisciplinary coauthorships place projects in better journals

Outline

Outline

Outline

Two Step Flow

- Paul Lazarsfeld, Bernard Berelson, and Hazel Gaudet write *The People's Choice* in 1944
- They find informal, personal contacts were mentioned far more frequently than exposure to radio or newspaper as sources of influence on voting behavior.
- Opinion leaders pass on their own interpretations in addition to the actual media content.
- Opinion leaders are quite influential in getting people to change their attitudes and behaviors and are quite similar to those they influence.

Two Step Flow

- Paul Lazarsfeld, Bernard Berelson, and Hazel Gaudet write *The People's Choice* in 1944
- They find informal, personal contacts were mentioned far more frequently than exposure to radio or newspaper as sources of influence on voting behavior.
- Opinion leaders pass on their own interpretations in addition to the actual media content.
- Opinion leaders are quite influential in getting people to change their attitudes and behaviors and are quite similar to those they influence.

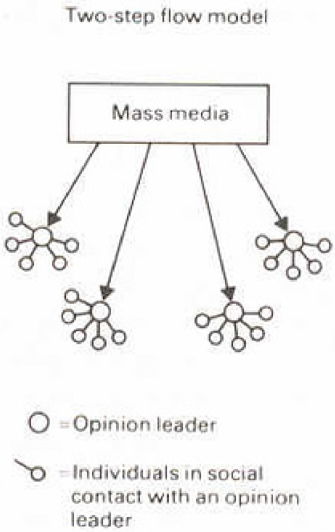
Two Step Flow

- Paul Lazarsfeld, Bernard Berelson, and Hazel Gaudet write *The People's Choice* in 1944
- They find informal, personal contacts were mentioned far more frequently than exposure to radio or newspaper as sources of influence on voting behavior.
- Opinion leaders pass on their own interpretations in addition to the actual media content.
- Opinion leaders are quite influential in getting people to change their attitudes and behaviors and are quite similar to those they influence.

Two Step Flow

- Paul Lazarsfeld, Bernard Berelson, and Hazel Gaudet write *The People's Choice* in 1944
- They find informal, personal contacts were mentioned far more frequently than exposure to radio or newspaper as sources of influence on voting behavior.
- Opinion leaders pass on their own interpretations in addition to the actual media content.
- Opinion leaders are quite influential in getting people to change their attitudes and behaviors and are quite similar to those they influence.

Two Step Flow



Yet, largely discredited in literature on media and communication

- One step flow where information comes directly from media
- Yet opinion leaders can moderate information (Troldahl 2001)
- And current social media provides a wide array of opportunities to observe two-step flow (i.e. on Twitter think HT, RT and on Facebook think about sharing a post)

Yet, largely discredited in literature on media and communication

- One step flow where information comes directly from media
- Yet opinion leaders can moderate information (Troldahl 2001)
- And current social media provides a wide array of opportunities to observe two-step flow (i.e. on Twitter think HT, RT and on Facebook think about sharing a post)

Yet, largely discredited in literature on media and communication

- One step flow where information comes directly from media
- Yet opinion leaders can moderate information (Troldahl 2001)
- And current social media provides a wide array of opportunities to observe two-step flow (i.e. on Twitter think HT, RT and on Facebook think about sharing a post)

Retraction followed by Brookman and Kalla's transgender persuasion study

- ""Maybe the thing to convey in your blogpost" Green told me "would be something to the effect that, just because the data don't exist to demonstrate the effectiveness of this method of changing minds, doesn't mean the hypothesis is false. And now the real work begins." He thinks the canvassers should go out and do the study again, for real this time." – Ira Glass, from TAL blogpost
- Just for fun: Retraction Summary from This American Life
- Just for fun: Case Study

Retraction followed by Brookman and Kalla's transgender persuasion study

- "“Maybe the thing to convey in your blogpost” Green told me “would be something to the effect that, just because the data don't exist to demonstrate the effectiveness of this method of changing minds, doesn't mean the hypothesis is false. And now the real work begins.” He thinks the canvassers should go out and do the study again, for real this time." – Ira Glass, from TAL blogpost
- Just for fun: Retraction Summary from This American Life
- Just for fun: Case Study

Retraction followed by Brookman and Kalla's transgender persuasion study

- ""Maybe the thing to convey in your blogpost" Green told me "would be something to the effect that, just because the data don't exist to demonstrate the effectiveness of this method of changing minds, doesn't mean the hypothesis is false. And now the real work begins." He thinks the canvassers should go out and do the study again, for real this time." – Ira Glass, from TAL blogpost
- Just for fun: Retraction Summary from This American Life
- Just for fun: Case Study

Social Ballots

- Democracy Live

Finding Super Influencers

- Voter Circles
- Magnify

Finding Super Influencers

- Voter Circles
- Magnify

How can we find these central nodes? What traits do they have?

- Privately-owned data (hard to get from Facebook?)
- Surveys
- Public records (does this work?)
- Network census (super expensive)

How can we find these central nodes? What traits do they have?

- Privately-owned data (hard to get from Facebook?)
- Surveys
- Public records (does this work?)
- Network census (super expensive)

How can we find these central nodes? What traits do they have?

- Privately-owned data (hard to get from Facebook?)
- Surveys
- Public records (does this work?)
- Network census (super expensive)

How can we find these central nodes? What traits do they have?

- Privately-owned data (hard to get from Facebook?)
- Surveys
- Public records (does this work?)
- Network census (super expensive)

Tell me About Your Network

- Interactions with other people (who did you have dinner with last night?)
- Social exchange with other people (who waters your plants or walks your dog when you're traveling?)
- Roles of other people (who is your best friend?)

Tell me About Your Network

- Interactions with other people (who did you have dinner with last night?)
- Social exchange with other people (who waters your plants or walks your dog when you're traveling?)
- Roles of other people (who is your best friend?)

Tell me About Your Network

- Interactions with other people (who did you have dinner with last night?)
- Social exchange with other people (who waters your plants or walks your dog when you're traveling?)
- Roles of other people (who is your best friend?)

Mesearch

- Next I'm going to give you a set of network batteries.
- Write down the answers on a sheet of paper as you go.
- Please answer each question.

Mesearch

- Next I'm going to give you a set of network batteries.
- Write down the answers on a sheet of paper as you go.
- Please answer each question.

Mesearch

- Next I'm going to give you a set of network batteries.
- Write down the answers on a sheet of paper as you go.
- Please answer each question.

Naming Your Network (Burt et al 2012)

- From time to time, most people discuss important matters with other people they trust. These important matters may be personal, professional or social. The people with whom they discuss these matters with may be family, work, friends or advisors.Please list the 5 people with whom you discussed matters most important to you (in person, telephone, e-mail etc) over the last 6 months.

- Think about the people with whom you spend your free time/leisure time.Over the last 6 months, who are the 5 people you have been with the most often for informal social activities (e.g., lunch, coffee, dinner, drinks, movies, sports, visits)? Please list the 5 people with whom you have been with the most for informal social activities over the last 6 months.

- Think about the people you talk to about matters pertaining to your profession or subspecialty – for example, people you talk with about new professional organizations,associations, research networks, journals, conferences, etc. Please list the 5 people with whom you most frequently/most commonly discussed matters pertaining to your professional development (in person, telephone, e-mail etc) over the last 6 months.

- Think about the people who have been most important to you in helping you succeed and/or advance professionally over the course of your professional career.Please list the 5 people who have been most important to you in helping you succeedand/or advance professionally over the course of your career

Overlap?

- Did you list five people for each of the four questions?
- Did you list the same or different people?
- Are the people on your list similar to you in terms of age, gender, race, or political affiliation?
- If you have a Facebook account, how many friends do you have on Facebook?
- If you have a Twitter account, how many followers do you have on Twitter?
- Do you vote in local elections?
- How many years have you lived in your neighborhood? Do you have children under the age of 18? Pets? Do you live with a spouse or partner?

Overlap?

- Did you list five people for each of the four questions?
- Did you list the same or different people?
- Are the people on your list similar to you in terms of age, gender, race, or political affiliation?
- If you have a Facebook account, how many friends do you have on Facebook?
- If you have a Twitter account, how many followers do you have on Twitter?
- Do you vote in local elections?
- How many years have you lived in your neighborhood? Do you have children under the age of 18? Pets? Do you live with a spouse or partner?

Overlap?

- Did you list five people for each of the four questions?
- Did you list the same or different people?
- Are the people on your list similar to you in terms of age, gender, race, or political affiliation?
- If you have a Facebook account, how many friends do you have on Facebook?
- If you have a Twitter account, how many followers do you have on Twitter?
- Do you vote in local elections?
- How many years have you lived in your neighborhood? Do you have children under the age of 18? Pets? Do you live with a spouse or partner?

Overlap?

- Did you list five people for each of the four questions?
- Did you list the same or different people?
- Are the people on your list similar to you in terms of age, gender, race, or political affiliation?
- If you have a Facebook account, how many friends do you have on Facebook?
- If you have a Twitter account, how many followers do you have on Twitter?
- Do you vote in local elections?
- How many years have you lived in your neighborhood? Do you have children under the age of 18? Pets? Do you live with a spouse or partner?

Overlap?

- Did you list five people for each of the four questions?
- Did you list the same or different people?
- Are the people on your list similar to you in terms of age, gender, race, or political affiliation?
- If you have a Facebook account, how many friends do you have on Facebook?
- If you have a Twitter account, how many followers do you have on Twitter?
- Do you vote in local elections?
- How many years have you lived in your neighborhood? Do you have children under the age of 18? Pets? Do you live with a spouse or partner?

Overlap?

- Did you list five people for each of the four questions?
- Did you list the same or different people?
- Are the people on your list similar to you in terms of age, gender, race, or political affiliation?
- If you have a Facebook account, how many friends do you have on Facebook?
- If you have a Twitter account, how many followers do you have on Twitter?
- Do you vote in local elections?
- How many years have you lived in your neighborhood? Do you have children under the age of 18? Pets? Do you live with a spouse or partner?

Overlap?

- Did you list five people for each of the four questions?
- Did you list the same or different people?
- Are the people on your list similar to you in terms of age, gender, race, or political affiliation?
- If you have a Facebook account, how many friends do you have on Facebook?
- If you have a Twitter account, how many followers do you have on Twitter?
- Do you vote in local elections?
- How many years have you lived in your neighborhood? Do you have children under the age of 18? Pets? Do you live with a spouse or partner?

Why should political scientists study networks?

We are all connected.

- Any two Americans are connected by six intermediaries (Milgram 1967).
- Any two unrelated Web pages are separated by only 19 links (Albert, Jeong and Barabasi 1999).
- All of the 19 hijackers for 9-11 were tied together using shared and available data (addresses, telephone numbers, frequent-flier numbers) and a disproportionate number network metrics converge on the leader (Mohamad Atta) (Krebs 2001).

Why should political scientists study networks?

We are all connected.

- Any two Americans are connected by six intermediaries (Milgram 1967).
- Any two unrelated Web pages are separated by only 19 links (Albert, Jeong and Barabasi 1999).
- All of the 19 hijackers for 9-11 were tied together using shared and available data (addresses, telephone numbers, frequent-flier numbers) and a disproportionate number network metrics converge on the leader (Mohamad Atta) (Krebs 2001).

Why should political scientists study networks?

We are all connected.

- Any two Americans are connected by six intermediaries (Milgram 1967).
- Any two unrelated Web pages are separated by only 19 links (Albert, Jeong and Barabasi 1999).
- All of the 19 hijackers for 9-11 were tied together using shared and available data (addresses, telephone numbers, frequent-flier numbers) and a disproportionate number network metrics converge on the leader (Mohamad Atta) (Krebs 2001).

Why should political scientists study networks?

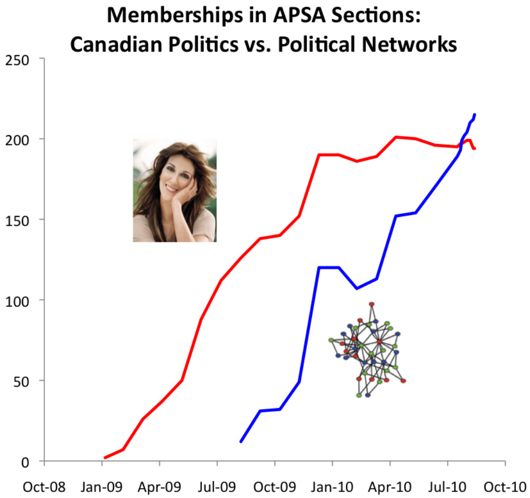
We are all connected.

- Any two Americans are connected by six intermediaries (Milgram 1967).
- Any two unrelated Web pages are separated by only 19 links (Albert, Jeong and Barabasi 1999).
- All of the 19 hijackers for 9-11 were tied together using shared and available data (addresses, telephone numbers, frequent-flier numbers) and a disproportionate number network metrics converge on the leader (Mohamad Atta) (Krebs 2001).

Why should political scientists study networks?

"We would probably all agree that one primary tie among political scientists is our emphasis on power, and understanding how and why power is used. We are all inherently interested in the exercise of power between and among individuals and groups and the implications that this exercise holds for social outcomes. We contend that this unifying concept is, at its very core, *relational* (McClurg and Young 2011)."

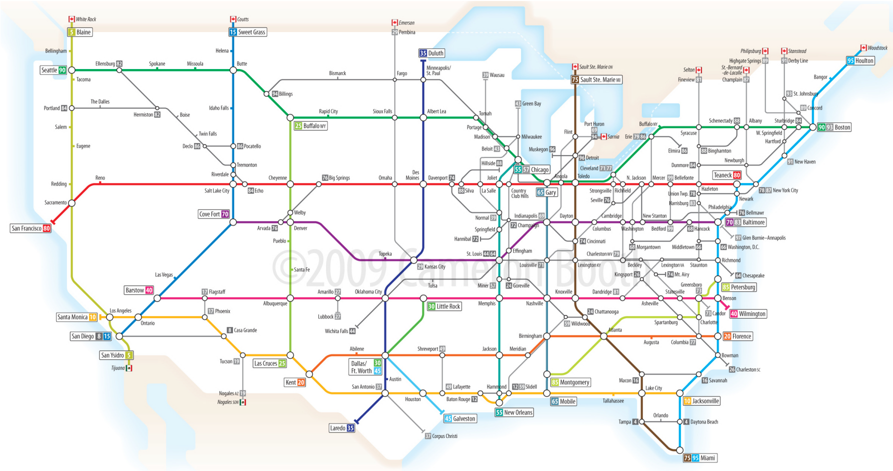
Why should political scientists study networks?



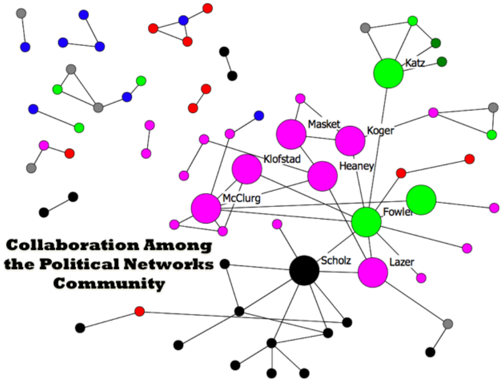
What is and is not a network?

- A set of actors and their relationships.

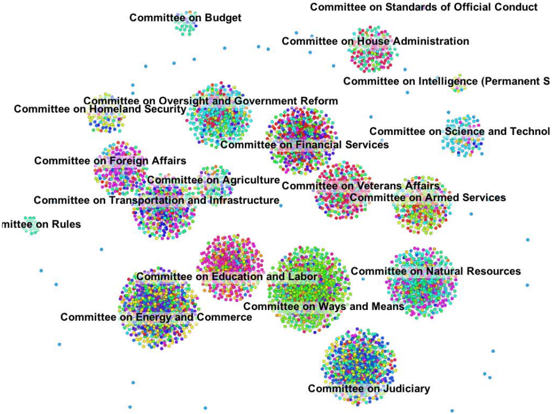
Example: Cities Connected by the Eisenhower Interstate System



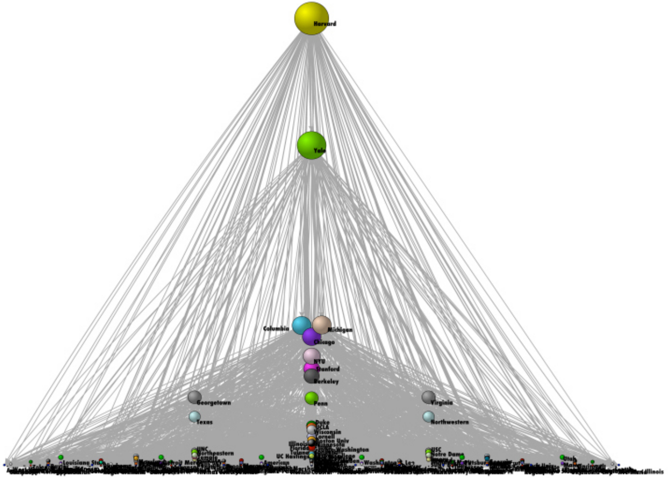
Example: Research Collaboration Within the Political Networks Conference Community (Berardo 2009)



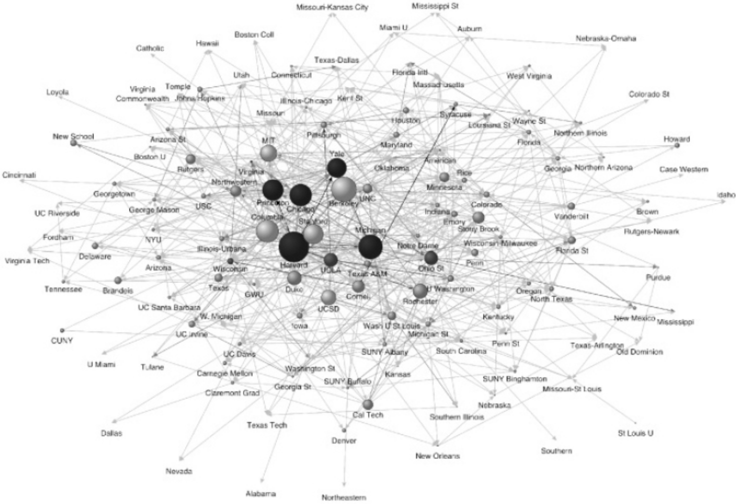
Example: All 6529 bills introduced in the 111th House of Representatives, computer-coded by topic and grouped by primary committee of reference (Gailmard and Patty 2013)



Example: Hiring Networks in Law vs Political Science (Katz et al. 2009)



Example: Hiring Networks in Law vs Political Science



What is and is not a network?

- A set of actors and their relationships.
- Nodes typically represent actors or institutions while edges represent connections between such entities.
- Edges can be either directed, to represent connections that flow from one node to another, or undirected.
- Edges can simply signify the existence of a connection, taking on a discrete value of 0 or 1, or they can be weighted to reflect the strength of the connection between two nodes.

What is and is not a network?

- A set of actors and their relationships.
- Nodes typically represent actors or institutions while edges represent connections between such entities.
- Edges can be either directed, to represent connections that flow from one node to another, or undirected.
- Edges can simply signify the existence of a connection, taking on a discrete value of 0 or 1, or they can be weighted to reflect the strength of the connection between two nodes.

What is and is not a network?

- A set of actors and their relationships.
- Nodes typically represent actors or institutions while edges represent connections between such entities.
- Edges can be either directed, to represent connections that flow from one node to another, or undirected.
- Edges can simply signify the existence of a connection, taking on a discrete value of 0 or 1, or they can be weighted to reflect the strength of the connection between two nodes.

What is and is not a network?

- A set of actors and their relationships.
- Nodes typically represent actors or institutions while edges represent connections between such entities.
- Edges can be either directed, to represent connections that flow from one node to another, or undirected.
- Edges can simply signify the existence of a connection, taking on a discrete value of 0 or 1, or they can be weighted to reflect the strength of the connection between two nodes.

What is and is not a network?

- Edges can represent a wide class of potential connections including social ties, the flow of goods between firms or countries, and the linkages between various actors or institutions.
  - Example: A *social network* is a social structure made up of a set of social actors (such as individuals or organizations) and a complex set of the dyadic ties between these actors (Wasserman and Faust 1994).
  - Example: A *political network* consists of the social network structure that focuses on politics, elections, or government (Sinclair 2012).
  - Just for fun: Seth Masket asks "But Is It A Network?" from his IGNITE Talk at Boulder PolNet 2012.

What is and is not a network?

- Edges can represent a wide class of potential connections including social ties, the flow of goods between firms or countries, and the linkages between various actors or institutions.
  - Example: A *social network* is a social structure made up of a set of social actors (such as individuals or organizations) and a complex set of the dyadic ties between these actors (Wasserman and Faust 1994).
  - Example: A *political network* consists of the social network structure that focuses on politics, elections, or government (Sinclair 2012).
  - Just for fun: Seth Masket asks "But Is It A Network?" from his IGNITE Talk at Boulder PolNet 2012.

What is and is not a network?

- Edges can represent a wide class of potential connections including social ties, the flow of goods between firms or countries, and the linkages between various actors or institutions.
  - Example: A *social network* is a social structure made up of a set of social actors (such as individuals or organizations) and a complex set of the dyadic ties between these actors (Wasserman and Faust 1994).
  - Example: A *political network* consists of the social network structure that focuses on politics, elections, or government (Sinclair 2012).
  - Just for fun: Seth Masket asks "But Is It A Network?" from his IGNITE Talk at Boulder PolNet 2012.

What is and is not a network?

- Edges can represent a wide class of potential connections including social ties, the flow of goods between firms or countries, and the linkages between various actors or institutions.
    - Example: A *social network* is a social structure made up of a set of social actors (such as individuals or organizations) and a complex set of the dyadic ties between these actors (Wasserman and Faust 1994).
    - Example: A *political network* consists of the social network structure that focuses on politics, elections, or government (Sinclair 2012).
    - Just for fun: Seth Masket asks "But Is It A Network?" from his IGNITE Talk at Boulder PolNet 2012.

Outline

Visualizing Contagion

- Plotting your data allows you to make sense of it (so long as it isn't too big).
- Just for fun: Retweeting Network Animation

Visualizing Contagion

- Plotting your data allows you to make sense of it (so long as it isn't too big).
- Just for fun: Retweeting Network Animation

Florentine Family Marriages in the early 15th century

- During the 1400s, Florence was ruled by an oligarchy of elite families.
- The Medici family rose to power in Florence, despite having less wealth and less political clout than other ruling families in Florence at that point in time.
- How can we understand their rise to power?

Florentine Family Marriages in the early 15th century

- During the 1400s, Florence was ruled by an oligarchy of elite families.
- The Medici family rose to power in Florence, despite having less wealth and less political clout than other ruling families in Florence at that point in time.
- How can we understand their rise to power?

Florentine Family Marriages in the early 15th century

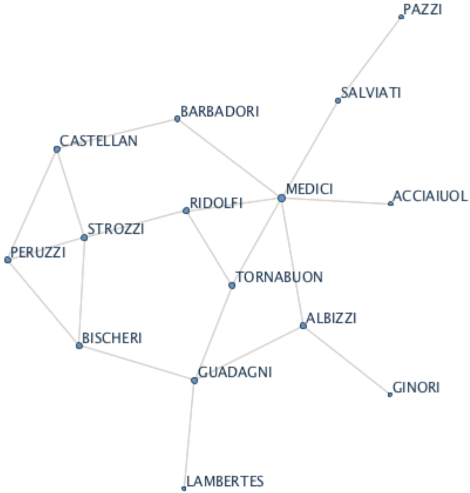
- During the 1400s, Florence was ruled by an oligarchy of elite families.
- The Medici family rose to power in Florence, despite having less wealth and less political clout than other ruling families in Florence at that point in time.
- How can we understand their rise to power?

Florentine Family Marriages in the early 15th century

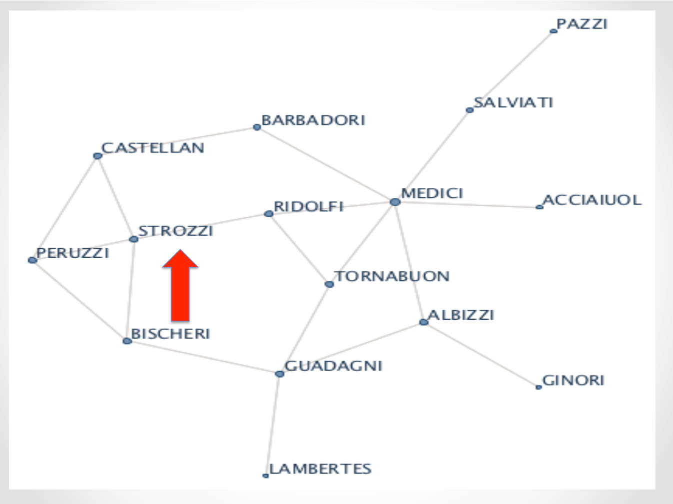
- During the 1400s, Florence was ruled by an oligarchy of elite families.
- The Medici family rose to power in Florence during the 1400s, despite having less wealth and less political clout than other ruling families in Florence at that point in time.
- How can we understand their rise to power?

Cosimo de' Medici "the godfather of the Renaissance" orchestrated strategic marriages to ensure a central position within the Florentine social network.

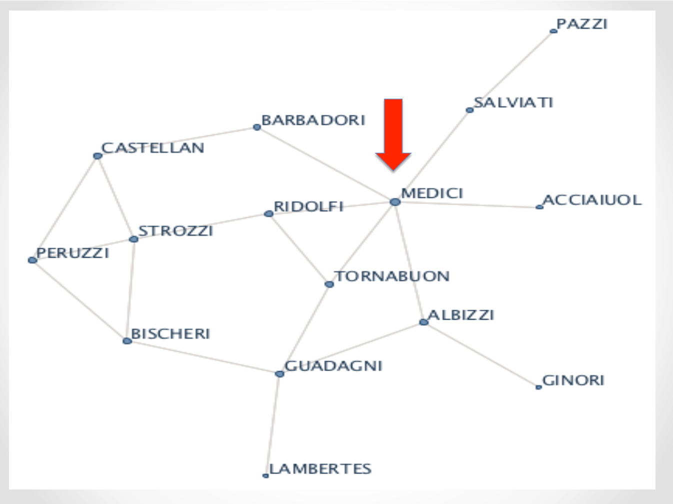
Example: Florentine Family Marriages in the early 15th century (Padgett and Ansell 1993)



The Strozzi Have the Most Wealth, But...



...the Medici Are Better Positioned



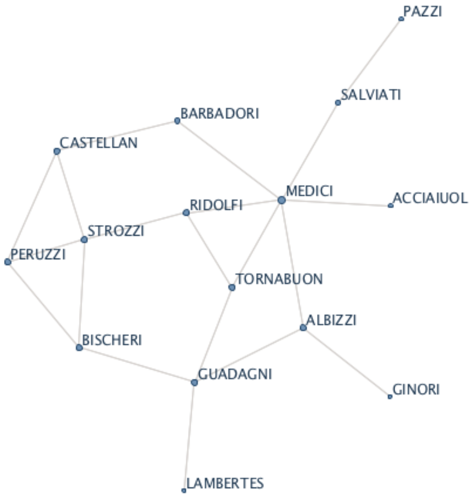
How Can We Quantify Network Position?

- Degree Centrality: Count the number of families a given family is linked to through marriages
- Betweenness Centrality: Calculate the fraction of the total number of shortest paths between any two families that a particular family lies

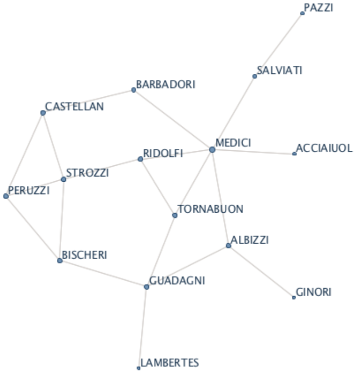
How Can We Quantify Network Position?

- Degree Centrality: Count the number of families a given family is linked to through marriages
- Betweenness Centrality: Calculate the fraction of the total number of shortest paths between any two families that a particular family lies

Degree Centrality: Medici = 6, Strozzi = 4

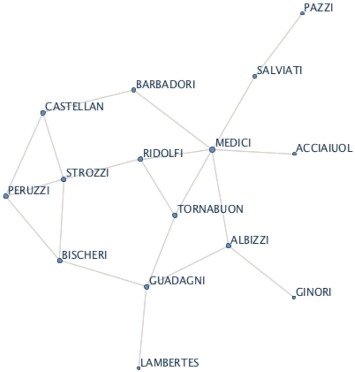


Betweenness Centrality: Medici = .522, Strozzi = .103



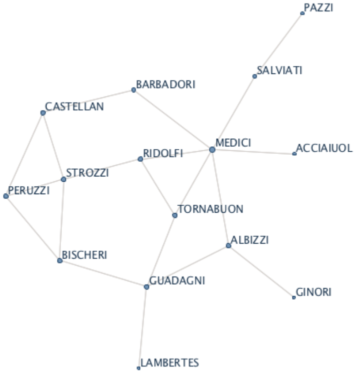
- Consider the shortest paths between the Barbadori and Guadagni families
- There are two such paths: Barbadori-Medici-Albizzi-Guadagni and Barbadori-Medici-Tornabuon-Guadagni
- The Medici lie on both paths and the Strozzi lie on neither

Betweenness Centrality: Medici = .522, Strozzi = .103



- Consider the shortest paths between the Barbadori and Guadagni families
- There are two such paths: Barbadori-Medici-Albizzi-Guadagni and Barbadori-Medici-Tornabuon-Guadagni
- The Medici lie on both paths and the Strozzi lie on neither

Betweenness Centrality: Medici = .522, Strozzi = .103



- Consider the shortest paths between the Barbadori and Guadagni families
- There are two such paths: Barbadori-Medici-Albizzi-Guadagni and Barbadori-Medici-Tornabuon-Guadagni
- The Medici lie on both paths and the Strozzi lie on neither

Quantifying Network Position

1. Degree: How connected a node is;
2. Closeness: How easily a node can reach other nodes;
3. Betweenness: How important a node is in terms of connecting other nodes; and
4. Neighbors' characteristics: How important, central, or influential a node's neighbors are.

Quantifying Network Position

1. Degree: How connected a node is;
2. Closeness: How easily a node can reach other nodes;
3. Betweenness: How important a node is in terms of connecting other nodes; and
4. Neighbors' characteristics: How important, central, or influential a node's neighbors are.

Quantifying Network Position

1. Degree: How connected a node is;
2. Closeness: How easily a node can reach other nodes;
3. Betweenness: How important a node is in terms of connecting other nodes; and
4. Neighbors' characteristics: How important, central, or influential a node's neighbors are.

Quantifying Network Position

1. Degree: How connected a node is;
2. Closeness: How easily a node can reach other nodes;
3. Betweenness: How important a node is in terms of connecting other nodes; and
4. Neighbors' characteristics: How important, central, or influential a node's neighbors are.

Quantifying Network Position: The Karate Club

- Observations from a karate club for 3 years (1970-1972) whose activities included social affairs (parties, dances, banquets, etc.) as well as regularly scheduled karate lessons.
- Club president (John A.) and part-time karate instructor (Mr Hi) disagreed over the price of karate lessons (and who had the authority to raise prices).
- Over time, the entire club became divided over this issue.
- After a series of confrontations, the supporters of Mr Hi resigned and formed a new organization.
- The structure of friendships within the club (those individuals who consistently interacted outside the normal activities of the club) provide insight into the split.
- Just for fun: Network Scientists and the Karate Club Award

Quantifying Network Position: The Karate Club

- Observations from a karate club for 3 years (1970-1972) whose activities included social affairs (parties, dances, banquets, etc.) as well as regularly scheduled karate lessons.
- Club president (John A.) and part-time karate instructor (Mr Hi) disagreed over the price of karate lessons (and who had the authority to raise prices).
- Over time, the entire club became divided over this issue.
- After a series of confrontations, the supporters of Mr Hi resigned and formed a new organization.
- The structure of friendships within the club (those individuals who consistently interacted outside the normal activities of the club) provide insight into the split.
- Just for fun: Network Scientists and the Karate Club Award

Quantifying Network Position: The Karate Club

- Observations from a karate club for 3 years (1970-1972) whose activities included social affairs (parties, dances, banquets, etc.) as well as regularly scheduled karate lessons.
- Club president (John A.) and part-time karate instructor (Mr Hi) disagreed over the price of karate lessons (and who had the authority to raise prices).
- Over time, the entire club became divided over this issue.
- After a series of confrontations, the supporters of Mr Hi resigned and formed a new organization.
- The structure of friendships within the club (those individuals who consistently interacted outside the normal activities of the club) provide insight into the split.
- Just for fun: Network Scientists and the Karate Club Award

Quantifying Network Position: The Karate Club

- Observations from a karate club for 3 years (1970-1972) whose activities included social affairs (parties, dances, banquets, etc.) as well as regularly scheduled karate lessons.
- Club president (John A.) and part-time karate instructor (Mr Hi) disagreed over the price of karate lessons (and who had the authority to raise prices).
- Over time, the entire club became divided over this issue.
- After a series of confrontations, the supporters of Mr Hi resigned and formed a new organization.
- The structure of friendships within the club (those individuals who consistently interacted outside the normal activities of the club) provide insight into the split.
- Just for fun: Network Scientists and the Karate Club Award

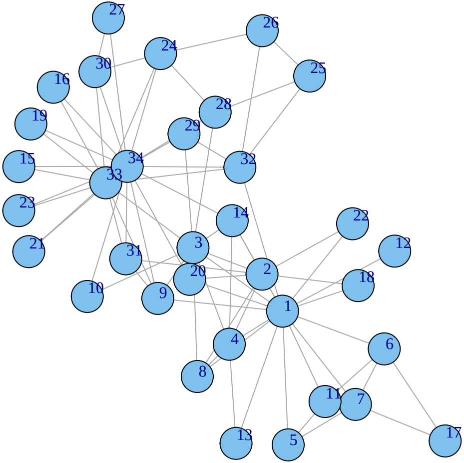
Quantifying Network Position: The Karate Club

- Observations from a karate club for 3 years (1970-1972) whose activities included social affairs (parties, dances, banquets, etc.) as well as regularly scheduled karate lessons.
- Club president (John A.) and part-time karate instructor (Mr Hi) disagreed over the price of karate lessons (and who had the authority to raise prices).
- Over time, the entire club became divided over this issue.
- After a series of confrontations, the supporters of Mr Hi resigned and formed a new organization.
- The structure of friendships within the club (those individuals who consistently interacted outside the normal activities of the club) provide insight into the split.
- Just for fun: Network Scientists and the Karate Club Award

Quantifying Network Position: The Karate Club

- Observations from a karate club for 3 years (1970-1972) whose activities included social affairs (parties, dances, banquets, etc.) as well as regularly scheduled karate lessons.
- Club president (John A.) and part-time karate instructor (Mr Hi) disagreed over the price of karate lessons (and who had the authority to raise prices).
- Over time, the entire club became divided over this issue.
- After a series of confrontations, the supporters of Mr Hi resigned and formed a new organization.
- The structure of friendships within the club (those individuals who consistently interacted outside the normal activities of the club) provide insight into the split.
- Just for fun: Network Scientists and the Karate Club Award

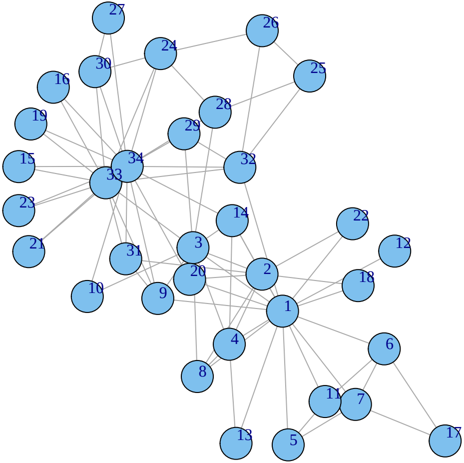
Quantifying Network Position: The Karate Club



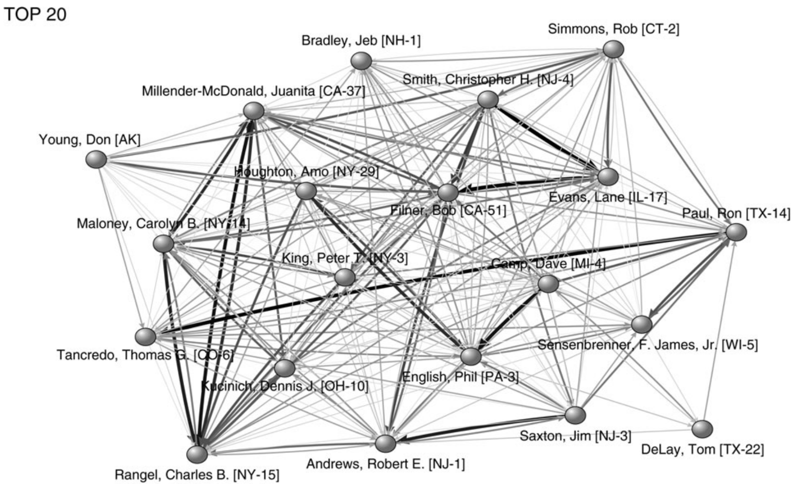
Quantifying Network Position: The Karate Club

Node	Degree	Closeness	Betweenness	Eigenvector
1	16	0.017	231.07	0.952
2	9	0.014	28.47	0.712
3	10	0.016	75.85	0.849
...	...	...	...	...
12	1	0.011	0.00	0.141

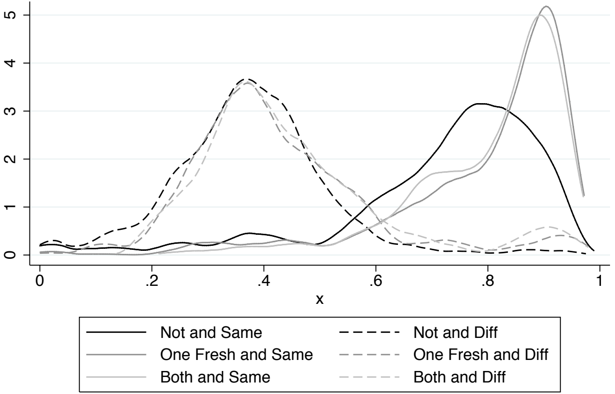
Quantifying Network Position: The Karate Club



Example: Legislative Centrality Cosponsorship (Fowler 2006)



Example: Legislative Centrality and the CA Blanket Primary (Alvarez and Sinclair 2011)



BUT...

- In these examples, we could draw pictures because we could see the network ties very explicitly. Marriages, retweets, karate club parties, cosponsorship and covoting are visible.
- What happens when you survey someone about their network? Do they say the right things? Can you draw the right pictures?

BUT...

- In these examples, we could draw pictures because we could see the network ties very explicitly. Marriages, retweets, karate club parties, cosponsorship and covoting are visible.
- What happens when you survey someone about their network? Do they say the right things? Can you draw the right pictures?

Outline

Uncovering Latent Social Structure

- Latent social structure of a network is often not fully observable and so needs to be constructed from what is observed.
- Questions that could be answered from uncovering these structures include: are there specific biases in a society, such as in hiring or publishing? Are there systematic ways to classify and categorize political ideologies or economic patterns of behavior?
- A *network community* is a subset of a social network graph that is more internally connected than externally connected.

Uncovering Latent Social Structure

- Latent social structure of a network is often not fully observable and so needs to be constructed from what is observed.
- Questions that could be answered from uncovering these structures include: are there specific biases in a society, such as in hiring or publishing? Are there systematic ways to classify and categorize political ideologies or economic patterns of behavior?
- A *network community* is a subset of a social network graph that is more internally connected than externally connected.

Uncovering Latent Social Structure

- Latent social structure of a network is often not fully observable and so needs to be constructed from what is observed.
- Questions that could be answered from uncovering these structures include: are there specific biases in a society, such as in hiring or publishing? Are there systematic ways to classify and categorize political ideologies or economic patterns of behavior?
- A *network community* is a subset of a social network graph that is more internally connected than externally connected.

Community Detection

- Community detection algorithms do not presuppose knowledge of the number and size of the underlying groups.
- Find non-overlapping communities.
- Rely upon *modularity maximization*, formalized by Girvan and Newman (2002).

Community Detection

- Community detection algorithms do not presuppose knowledge of the number and size of the underlying groups.
- Find non-overlapping communities.
- Rely upon *modularity maximization*, formalized by Girvan and Newman (2002).

Community Detection

- Community detection algorithms do not presuppose knowledge of the number and size of the underlying groups.
- Find non-overlapping communities.
- Rely upon *modularity maximization*, formalized by Girvan and Newman (2002).

Modularity

Let $\mathbf{e}$ be a symmetric $n \times n$ matrix where each element e_{ij} represents the fraction of all edges that link community i to vertices in community j. Modularity is defined, then, as:

$$Q = \sum_i (e_{ii} - a_i^2) = \mathrm{Tr}\ \mathbf{e} - ||\mathbf{e}^2||$$

where a_i is defined as the row sums, $a_i = \sum_j e_{ij}$. Tr $\mathbf{e}$, the trace of $\mathbf{e}$ gives the sum of edges connecting vertices in the same community, and $||\mathbf{e}^2||$ represents the sum of the elements of the matrix $\mathbf{e}^2$.

- We can think of $||\mathbf{e}^2||$ as measuring, holding the detected community structure fixed, the expected number of edges connecting communities if connections between vertices were random.
- Modularity, Q, can then take on values between 0 and 1.

Modularity

Let $\mathbf{e}$ be a symmetric $n \times n$ matrix where each element e_{ij} represents the fraction of all edges that link community i to vertices in community j. Modularity is defined, then, as:

$$Q = \sum_i (e_{ii} - a_i^2) = \text{Tr } \mathbf{e} - ||\mathbf{e}^2||$$

where a_i is defined as the row sums, $a_i = \sum_j e_{ij}$. Tr $\mathbf{e}$, the trace of $\mathbf{e}$ gives the sum of edges connecting vertices in the same community, and $||\mathbf{e}^2||$ represents the sum of the elements of the matrix $\mathbf{e}^2$.

- We can think of $||\mathbf{e}^2||$ as measuring, holding the detected community structure fixed, the expected number of edges connecting communities if connections between vertices were random.
- Modularity, Q, can then take on values between 0 and 1.

Modularity

Let $\mathbf{e}$ be a symmetric $n \times n$ matrix where each element e_{ij} represents the fraction of all edges that link community i to vertices in community j. Modularity is defined, then, as:

$$Q = \sum_i (e_{ii} - a_i^2) = \text{Tr } \mathbf{e} - ||\mathbf{e}^2||$$

where a_i is defined as the row sums, $a_i = \sum_j e_{ij}$. Tr $\mathbf{e}$, the trace of $\mathbf{e}$ gives the sum of edges connecting vertices in the same community, and $||\mathbf{e}^2||$ represents the sum of the elements of the matrix $\mathbf{e}^2$.

- We can think of $||\mathbf{e}^2||$ as measuring, holding the detected community structure fixed, the expected number of edges connecting communities if connections between vertices were random.
- Modularity, Q, can then take on values between 0 and 1.

Steps of Modularity Maximization

1. Calculate betweenness (the number of shortest paths that run along a given edge) for every edge.
2. The edge with the highest betweenness score is removed, as this edge is most likely to connect communities as opposed to lie within communities.
3. Estimate modularity (measures the number of within-community edges relative to a null model of a random graph with the same degree distribution).
4. The betweenness scores are recalculated within the new graph, and the process repeats.
5. Continue to remove edges until modularity is maximized.

Steps of Modularity Maximization

1. Calculate betweenness (the number of shortest paths that run along a given edge) for every edge.
2. The edge with the highest betweenness score is removed, as this edge is most likely to connect communities as opposed to lie within communities.
3. Estimate modularity (measures the number of within-community edges relative to a null model of a random graph with the same degree distribution).
4. The betweenness scores are recalculated within the new graph, and the process repeats.
5. Continue to remove edges until modularity is maximized.

Steps of Modularity Maximization

1. Calculate betweenness (the number of shortest paths that run along a given edge) for every edge.
2. The edge with the highest betweenness score is removed, as this edge is most likely to connect communities as opposed to lie within communities.
3. Estimate modularity (measures the number of within-community edges relative to a null model of a random graph with the same degree distribution).
4. The betweenness scores are recalculated within the new graph, and the process repeats.
5. Continue to remove edges until modularity is maximized.

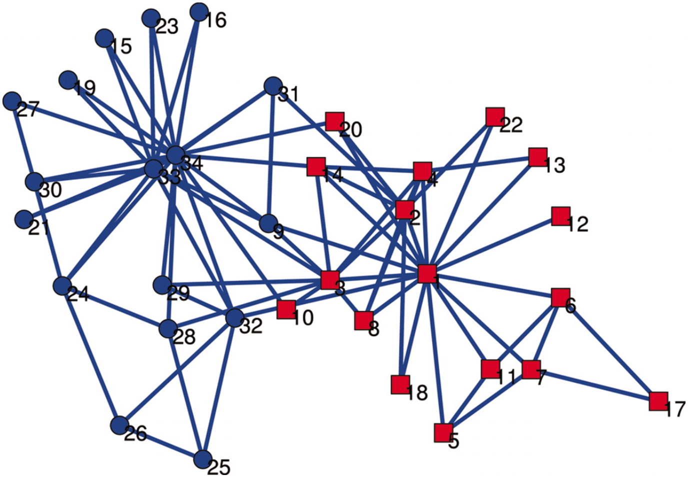
Steps of Modularity Maximization

1. Calculate betweenness (the number of shortest paths that run along a given edge) for every edge.
2. The edge with the highest betweenness score is removed, as this edge is most likely to connect communities as opposed to lie within communities.
3. Estimate modularity (measures the number of within-community edges relative to a null model of a random graph with the same degree distribution).
4. The betweenness scores are recalculated within the new graph, and the process repeats.
5. Continue to remove edges until modularity is maximized.

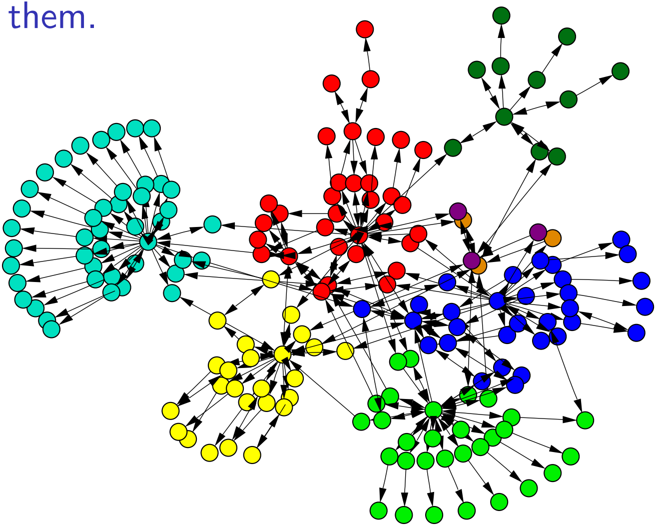
Steps of Modularity Maximization

1. Calculate betweenness (the number of shortest paths that run along a given edge) for every edge.
2. The edge with the highest betweenness score is removed, as this edge is most likely to connect communities as opposed to lie within communities.
3. Estimate modularity (measures the number of within-community edges relative to a null model of a random graph with the same degree distribution).
4. The betweenness scores are recalculated within the new graph, and the process repeats.
5. Continue to remove edges until modularity is maximized.

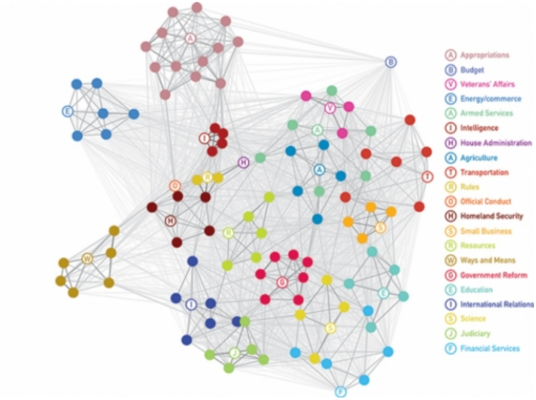
An example: The Karate Club

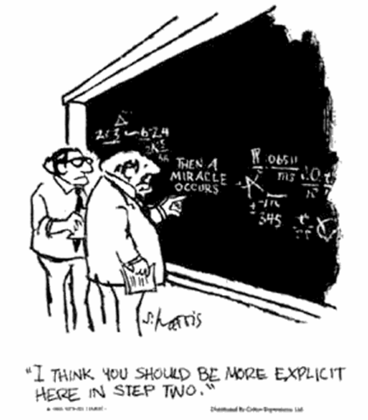


An example: Pages on a web site and the hyperlinks between them.

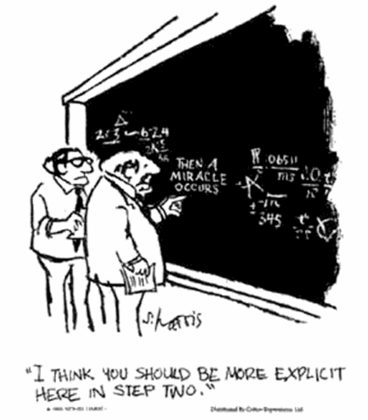


An example: Congressional Roll Call Voting and Committees



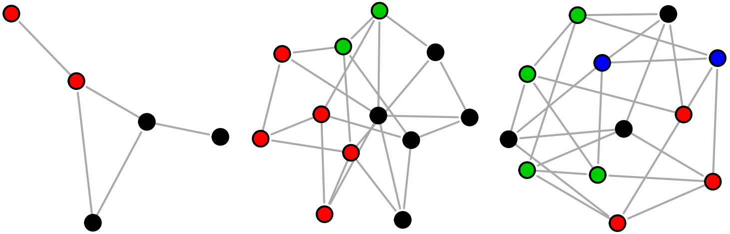


- In the field of network analysis there are many possible algorithms to detect *communities* within observed networks.
- A cautionary note: much of the literature takes an “inter-occular test” approach



- In the field of network analysis there are many possible algorithms to detect *communities* within observed networks.
- A cautionary note: much of the literature takes an "inter-occular test" approach

An example of a graph with 2 communities, 3 communities and 4 communities.



Lifestyles of Democrats and Republicans

- Can we find a community in survey data?
- Are there systematic and durable differences in lifestyles that arise not only from the standard cleavages in American life but rather from partisanship?
- Can community detection algorithms find other kinds of latent communities?

Lifestyles of Democrats and Republicans

- Can we find a community in survey data?
- Are there systematic and durable differences in lifestyles that arise not only from the standard cleavages in American life but rather from partisanship?
- Can community detection algorithms find other kinds of latent communities?

Lifestyles of Democrats and Republicans

- Can we find a community in survey data?
- Are there systematic and durable differences in lifestyles that arise not only from the standard cleavages in American life but rather from partisanship?
- Can community detection algorithms find other kinds of latent communities?

Is a latte-drinking dog owner more likely to be a Democrat?

- We utilize the American Panel Survey (TAPS) which solicits data from 1,400 respondents on 300 of their lifestyle choices.
- We employ a community detection algorithm (don't specify the number of communities you are looking for) and a conventional latent-class analysis (do specify the number of latent classes you are looking for).
- We control for the conventional explanations: gender, race, age, income, education, region.

Is a latte-drinking dog owner more likely to be a Democrat?

- We utilize the American Panel Survey (TAPS) which solicits data from 1,400 respondents on 300 of their lifestyle choices.
- We employ a community detection algorithm (don't specify the number of communities you are looking for) and a conventional latent-class analysis (do specify the number of latent classes you are looking for).
- We control for the conventional explanations: gender, race, age, income, education, region.

Is a latte-drinking dog owner more likely to be a Democrat?

- We utilize the American Panel Survey (TAPS) which solicits data from 1,400 respondents on 300 of their lifestyle choices.
- We employ a community detection algorithm (don't specify the number of communities you are looking for) and a conventional latent-class analysis (do specify the number of latent classes you are looking for).
- We control for the conventional explanations: gender, race, age, income, education, region.

Table 1: Categories of Non-political Variables

Television	Fast Food	Cars
Sports Watching	Activities	National Parks
Super Heroes	Super Powers	Coffee
Food	Living Space	Non-Political Figures
Public Services	Desserts	Healthy Living
Smoking	Music Tastes	Musical Abilities
Sense of Humor	Pets	Reading
Exercise	Daily Habits	Communication
Online Time	Hobbies	

Table 2: Comparing Classifications

	Community=0	Community=1	Total
Latent Class=0	618	141	759
Latent Class=1	75	566	641
Total	693	707	1400

Note: Rows 1 & 2 and Columns 1 & 2 report cross-classification results for the dichotomous Latent Class (LCA) and Community Detection (CDA) analyses.

Table 3: Predicting Classification

	LCA	CDA
(Intercept)	−3.37*	−3.59*
	(0.64)	(0.61)
7-Point Party ID	0.11*	0.06*
	(0.03)	(0.03)
MSA Category	−0.12	0.22
	(0.17)	(0.16)
White	−0.05	0.09
	(0.16)	(0.15)
Midwest	−0.25	−0.18
	(0.20)	(0.19)
South	−0.04	−0.18
	(0.19)	(0.18)
West	−0.08	0.07
	(0.20)	(0.19)
Political Interest	0.15*	0.02
	(0.08)	(0.07)
Age	0.06*	0.04*
	(0.00)	(0.00)
Female	0.27*	0.34*
	(0.12)	(0.12)
Education	−0.02	0.10*
	(0.04)	(0.03)
Income	−0.09*	−0.06*
	(0.02)	(0.02)
N	1344	1344

Extremists

- Partisan extremists: people who report "Strong Democrat" or "Strong Republican" on the 7-point PID scale
- Lifestyle extremists: PCA over 300 lifestyle items to establish a scale, pull individuals in the top and bottom deciles of the scores distribution

Extremists

- Partisan extremists: people who report "Strong Democrat" or "Strong Republican" on the 7-point PID scale
- Lifestyle extremists: PCA over 300 lifestyle items to establish a scale, pull individuals in the top and bottom deciles of the scores distribution

Table 4: Cross-partisan deliberation and discussion

	Outcome variable:
	Days of the week respondent discusses politics
Lifestyle extreme	−0.374*
	(0.168)
PID extreme	0.573*
	(0.125)
PID extreme	0.358
× Lifestyle extreme	(0.288)
Constant	2.314*
	(0.074)
Observations	1,328
R^2	0.026
Adjusted R^2	0.024
Residual Std. Error	1.974 (df = 1324)

Note: Standard errors in parentheses. $^*p < 0.05$

[htp]

Outline

Survey Research: Descriptive Inference

- Most survey research on networks is not concerned about the structural relationship across survey respondents but instead each individual respondents' political environment (contextual or personal)
- That is, a significant body of literature has been created around linking name generators to political behaviors and public opinion
- This places most of network survey research in the *descriptive inference* framework: surveys are being used to characterize the key features of the respondents' lives (but not to make predictions or draw causal inferences)

Survey Research: Descriptive Inference

- Most survey research on networks is not concerned about the structural relationship across survey respondents but instead each individual respondents' political environment (contextual or personal)
- That is, a significant body of literature has been created around linking name generators to political behaviors and public opinion
- This places most of network survey research in the *descriptive inference* framework: surveys are being used to characterize the key features of the respondents' lives (but not to make predictions or draw causal inferences)

Survey Research: Descriptive Inference

- Most survey research on networks is not concerned about the structural relationship across survey respondents but instead each individual respondents' political environment (contextual or personal)
- That is, a significant body of literature has been created around linking name generators to political behaviors and public opinion
- This places most of network survey research in the *descriptive inference* framework: surveys are being used to characterize the key features of the respondents' lives (but not to make predictions or draw causal inferences)

Conventional Network Surveys

- The South Bend Study (Huckfeldt and Sprague 1984): Third wave of South Bend, Indiana survey which includes a name-generator and validated vote data.
- ANES 2006 (Sinclair 2012) asks network composition questions (as does the CCES 2008) via a name-generator. When compared with respect to partisan identification or candidate choices – network diversity affects how a respondent answers!

Conventional Network Surveys

- The South Bend Study (Huckfeldt and Sprague 1984): Third wave of South Bend, Indiana survey which includes a name-generator and validated vote data.
- ANES 2006 (Sinclair 2012) asks network composition questions (as does the CCES 2008) via a name-generator. When compared with respect to partisan identification or candidate choices – network diversity affects how a respondent answers!

Generating network data via batteries other than name generators

- "position generator" or "resource generator"
- Toronto Social Support

Generating network data via batteries other than name generators

- "position generator" or "resource generator"
- Toronto Social Support

Network size: gleaned from surveys

- The size of an individual's online social network is measured along a six-point scale according to resondents' estimates of how many "people are on your list of 'followers', 'friends', 'connections', or 'contacts'" on the social networking site (SNS) that they use most often (Kahne and Bowyer 2018).
- Are online social networks a subset of name-generator networks?

Network size: gleaned from surveys

- The size of an individual's online social network is measured along a six-point scale according to resondents' estimates of how many "people are on your list of 'followers', 'friends', 'connections', or 'contacts'" on the social networking site (SNS) that they use most often (Kahne and Bowyer 2018).
- Are online social networks a subset of name-generator networks?

How many people do you know in prison? (Zheng et al 2006)

- How many X's do you know? (fill in X for "people do you know in prison" or "dentists" or "people whose first name is Betsy")
- For the prison question, the mean of the response for Americans is 1.0 because of the variability: almost 3% of Americans know at least 10 prisoners while 70% of the respondents report knowing 0.
- Asking questions like this allows us to use overdispersion that measures the variation in the relative propensities of individuals to form ties to a given social group

How many people do you know in prison? (Zheng et al 2006)

- How many X's do you know? (fill in X for "people do you know in prison" or "dentists" or "people whose first name is Betsy")
- For the prison question, the mean of the response for Americans is 1.0 because of the variability: almost 3% of Americans know at least 10 prisoners while 70% of the respondents report knowing 0.
- Asking questions like this allows us to use overdispersion that measures the variation in the relative propensities of individuals to form ties to a given social group

How many people do you know in prison? (Zheng et al 2006)

- How many X's do you know? (fill in X for "people do you know in prison" or "dentists" or "people whose first name is Betsy")
- For the prison question, the mean of the response for Americans is 1.0 because of the variability: almost 3% of Americans know at least 10 prisoners while 70% of the respondents report knowing 0.
- Asking questions like this allows us to use overdispersion that measures the variation in the relative propensities of individuals to form ties to a given social group

How many people do you know in group X?

- Respondents were told, "For the purposes of this study, the definition of knowing someone is that you know them and they know you by sight or by name, that you could contact them, that they live within the United States, and that there has been some contact (either in person, by telephone or mail) in the past 2 years."
- If you assume that everyone is equally-likely to know someone in the national population (a heroic assumption) and you know a fixed rate of the population (such as the number of people named Nicole), you can infer a guess as to the size of the respondent's network
- By comparing the dispersion of the responses (people named Nicole vs people in prison) we have a sense for how likely it is for people to form ties to specific groups

How many people do you know in group X?

- Respondents were told, "For the purposes of this study, the definition of knowing someone is that you know them and they know you by sight or by name, that you could contact them, that they live within the United States, and that there has been some contact (either in person, by telephone or mail) in the past 2 years."
- If you assume that everyone is equally-likely to know someone in the national population (a heroic assumption) and you know a fixed rate of the population (such as the number of people named Nicole), you can infer a guess as to the size of the respondent's network
- By comparing the dispersion of the responses (people named Nicole vs people in prison) we have a sense for how likely it is for people to form ties to specific groups

How many people do you know in group X?

- Respondents were told, "For the purposes of this study, the definition of knowing someone is that you know them and they know you by sight or by name, that you could contact them, that they live within the United States, and that there has been some contact (either in person, by telephone or mail) in the past 2 years."
- If you assume that everyone is equally-likely to know someone in the national population (a heroic assumption) and you know a fixed rate of the population (such as the number of people named Nicole), you can infer a guess as to the size of the respondent's network
- By comparing the dispersion of the responses (people named Nicole vs people in prison) we have a sense for how likely it is for people to form ties to specific groups

Survey Research: Surveys of Connected Respondents

- Surveys where respondents are drawn from a network – and are themselves connected – can be extremely problematic because of overlap in respondent networks.
- "Virtually all college students reported that their friends drank more than they did."
- "Virtually all network scientists reported their friends know more R than they do."
- Just for fun: Network simulation

Survey Research: Surveys of Connected Respondents

- Surveys where respondents are drawn from a network – and are themselves connected – can be extremely problematic because of overlap in respondent networks.
- "Virtually all college students reported that their friends drank more than they did."
- "Virtually all network scientists reported their friends know more R than they do."
- Just for fun: Network simulation

Survey Research: Surveys of Connected Respondents

- Surveys where respondents are drawn from a network – and are themselves connected – can be extremely problematic because of overlap in respondent networks.
- "Virtually all college students reported that their friends drank more than they did."
- "Virtually all network scientists reported their friends know more R than they do."
- Just for fun: Network simulation

Survey Research: Surveys of Connected Respondents

- Surveys where respondents are drawn from a network – and are themselves connected – can be extremely problematic because of overlap in respondent networks.
- "Virtually all college students reported that their friends drank more than they did."
- "Virtually all network scientists reported their friends know more R than they do."
- Just for fun: Network simulation

Snowball Sampling

- We sample differently when looking at respondents who are drawn from within a network. Indeed, there is a vast and vibrant literature on "respondent driven sampling" (RDS) which is sometimes called "snowball sampling"
- It is typically employed for hard-to-sample populations or sometimes "hidden" populations

Snowball Sampling

- We sample differently when looking at respondents who are drawn from within a network. Indeed, there is a vast and vibrant literature on "respondent driven sampling" (RDS) which is sometimes called "snowball sampling"
- It is typically employed for hard-to-sample populations or sometimes "hidden" populations

Snowball Sampling

- We know the complete network from a handful of network census projects (Project 90, which is a network census of high-risk heterosexuals in Colorado Springs (giant component has 4430 individuals and 18,407 edges that represent links for drug/sexual partners), and the National Longitudinal Study of Adolescent Health (mapped the friendship networks of 84 middle and high schools in the US, with 76,262 individuals and 258,688 edges))
- Modeling recruitment with a fairly reasonable set of assumptions (Goel and Salganik 2010, PNAS) it is plausible to believe that RDS can produce an unbiased estimate of the population trait of interest but the variability is much higher than typically acknowledged

Snowball Sampling

- We know the complete network from a handful of network census projects (Project 90, which is a network census of high-risk heterosexuals in Colorado Springs (giant component has 4430 individuals and 18,407 edges that represent links for drug/sexual partners), and the National Longitudinal Study of Adolescent Health (mapped the friendship networks of 84 middle and high schools in the US, with 76,262 individuals and 258,688 edges))
- Modeling recruitment with a fairly reasonable set of assumptions (Goel and Salganik 2010, PNAS) it is plausible to believe that RDS can produce an unbiased estimate of the population trait of interest but the variability is much higher than typically acknowledged

Outline

James Fowler: Connected

- Do health outcomes spread through networks?
- Problem: people select their networks (families, friends) because of shared traits (proclivities towards particular health outcomes, geography, etc.) that are causally associated with these health outcomes
- Classic omitted variable bias problem

James Fowler: Connected

- Do health outcomes spread through networks?
- Problem: people select their networks (families, friends) because of shared traits (proclivities towards particular health outcomes, geography, etc.) that are causally associated with these health outcomes
- Classic omitted variable bias problem

James Fowler: Connected

- Do health outcomes spread through networks?
- Problem: people select their networks (families, friends) because of shared traits (proclivities towards particular health outcomes, geography, etc.) that are causally associated with these health outcomes
- Classic omitted variable bias problem

James Fowler: Connected

- Do health outcomes spread through networks?
- Problem: people select their networks (families, friends) because of shared traits (proclivities towards particular health outcomes, geography, etc.) that are causally associated with these health outcomes
- Classic omitted variable bias problem

James Fowler: Connected

- Do health outcomes spread through networks?
- Problem: people select their networks (families, friends) because of shared traits (proclivities towards particular health outcomes, geography, etc.) that are causally associated with these health outcomes
- Classic omitted variable bias problem

James Fowler: Connected

- Do health outcomes spread through networks?
- Problem: people select their networks (families, friends) because of shared traits (proclivities towards particular health outcomes, geography, etc.) that are causally associated with these health outcomes
- Classic omitted variable bias problem

The Spread of Obesity in a Large Social Network over 32 Years

To the Animation!

Obesity in Networks

Fowler on Colbert

Data

- 12,067 people (5125 "egos" and the remainder "alters") from the Framingham Heart Study
- 1971-2003
- Physical exams (including weight) measured at regular intervals
- Social ties identified by ego, many of whom are also in the study (and hence provides longitudinal data)

Data

- 12,067 people (5125 "egos" and the remainder "alters") from the Framingham Heart Study
- 1971-2003
- Physical exams (including weight) measured at regular intervals
- Social ties identified by ego, many of whom are also in the study (and hence provides longitudinal data)

Data

- 12,067 people (5125 "egos" and the remainder "alters") from the Framingham Heart Study
- 1971-2003
- Physical exams (including weight) measured at regular intervals
- Social ties identified by ego, many of whom are also in the study (and hence provides longitudinal data)

Data

- 12,067 people (5125 "egos" and the remainder "alters") from the Framingham Heart Study
- 1971-2003
- Physical exams (including weight) measured at regular intervals
- Social ties identified by ego, many of whom are also in the study (and hence provides longitudinal data)

Model

Obesity (t+1) = Obesity (t) + alter's obesity (t and t+1) + geographic distance between alter and ego + ego attributes (age, sex, education)

3 Possible Explanations

1. Homophily (egos associate with like alters)
2. Confounding (egos and alters share unobserved attributes or experiences)
3. Induction (social influence)

3 Possible Explanations

1. Homophily (egos associate with like alters)
2. Confounding (egos and alters share unobserved attributes or experiences)
3. Induction (social influence)

3 Possible Explanations

1. Homophily (egos associate with like alters)
2. Confounding (egos and alters share unobserved attributes or experiences)
3. Induction (social influence)

Observational Data

Unfortunately, the field has largely concluded that observational data has very limited capabilities to draw causal inferences without strong assumptions

- "If your friend Joey jumped off a bridge, would you jump too?" (Shalizi and Thomas 2011)
- Yes: Joey inspires you (Induction).
- Yes: you're friends because you both like to jump off of bridges (Homophily).
- Yes: you're friends because you both have a common risk-seeking propensity (Confounding).
- Summary: These three explanations are (almost) impossible to distinguish in observational social network studies.

Observational Data

Unfortunately, the field has largely concluded that observational data has very limited capabilities to draw causal inferences without strong assumptions

- "If your friend Joey jumped off a bridge, would you jump too?" (Shalizi and Thomas 2011)
- Yes: Joey inspires you (Induction).
- Yes: you're friends because you both like to jump off of bridges (Homophily).
- Yes: you're friends because you both have a common risk-seeking propensity (Confounding).
- Summary: These three explanations are (almost) impossible to distinguish in observational social network studies.

Observational Data

Unfortunately, the field has largely concluded that observational data has very limited capabilities to draw causal inferences without strong assumptions

- "If your friend Joey jumped off a bridge, would you jump too?" (Shalizi and Thomas 2011)
- Yes: Joey inspires you (Induction).
- Yes: you're friends because you both like to jump off of bridges (Homophily).
- Yes: you're friends because you both have a common risk-seeking propensity (Confounding).
- Summary: These three explanations are (almost) impossible to distinguish in observational social network studies.

Observational Data

Unfortunately, the field has largely concluded that observational data has very limited capabilities to draw causal inferences without strong assumptions

- "If your friend Joey jumped off a bridge, would you jump too?" (Shalizi and Thomas 2011)
- Yes: Joey inspires you (Induction).
- Yes: you're friends because you both like to jump off of bridges (Homophily).
- Yes: you're friends because you both have a common risk-seeking propensity (Confounding).
- Summary: These three explanations are (almost) impossible to distinguish in observational social network studies.

Observational Data

Unfortunately, the field has largely concluded that observational data has very limited capabilities to draw causal inferences without strong assumptions

- "If your friend Joey jumped off a bridge, would you jump too?" (Shalizi and Thomas 2011)
- Yes: Joey inspires you (Induction).
- Yes: you're friends because you both like to jump off of bridges (Homophily).
- Yes: you're friends because you both have a common risk-seeking propensity (Confounding).
- Summary: These three explanations are (almost) impossible to distinguish in observational social network studies.

Room for Encouragement

- VanderWeele (2011) proposed a sensitivity analysis technique to assess the extent to which an unmeasured factor responsible for homophily or confounding would have to be related to both the ego's and the alter's state in order to substantially alter qualitative and quantitative conclusions.
- VanderWeele, Ogburn and Tchetgen Tchetgen (2012) indicate that there are reasonable conditions to use the kinds of models Fowler and Christakis advance.
- Substantively, it appears the Fowler and Christakis results are fairly robust to this skepticism, but there emerged a significant push to study the effects of networks experimentally.

Room for Encouragement

- VanderWeele (2011) proposed a sensitivity analysis technique to assess the extent to which an unmeasured factor responsible for homophily or confounding would have to be related to both the ego's and the alter's state in order to substantially alter qualitative and quantitative conclusions.
- VanderWeele, Ogburn and Tchetgen Tchetgen (2012) indicate that there are reasonable conditions to use the kinds of models Fowler and Christakis advance.
- Substantively, it appears the Fowler and Christakis results are fairly robust to this skepticism, but there emerged a significant push to study the effects of networks experimentally.

Room for Encouragement

- VanderWeele (2011) proposed a sensitivity analysis technique to assess the extent to which an unmeasured factor responsible for homophily or confounding would have to be related to both the ego's and the alter's state in order to substantially alter qualitative and quantitative conclusions.
- VanderWeele, Ogburn and Tchetgen Tchetgen (2012) indicate that there are reasonable conditions to use the kinds of models Fowler and Christakis advance.
- Substantively, it appears the Fowler and Christakis results are fairly robust to this skepticism, but there emerged a significant push to study the effects of networks experimentally.

Outline

Experimental Options

1. Randomly assign some members of the network to receive a stimulus and see how that percolates through the network.
2. Randomly assign a new network structure (new social ties, for example).

Experimental Options

1. Randomly assign some members of the network to receive a stimulus and see how that percolates through the network.
2. Randomly assign a new network structure (new social ties, for example).

The key is that you need something that is truly exogenous.

- "No causation without manipulation" (Holland 1986)
- You cannot estimate causal effects of immutable characteristics (gender, race, age, etc.)
- What does the causal effect of gender mean? Problem: CONFOUNDING
- For example, consider whether you can estimate the causal effect of having a discussion leader with certain preferences on deliberation outcomes (Humphreys et al 2006).
- Are networks immutable?

The key is that you need something that is truly exogenous.

- "No causation without manipulation" (Holland 1986)
- You cannot estimate causal effects of immutable characteristics (gender, race, age, etc.)
- What does the causal effect of gender mean? Problem: CONFOUNDING
- For example, consider whether you can estimate the causal effect of having a discussion leader with certain preferences on deliberation outcomes (Humphreys et al 2006).
- Are networks immutable?

The key is that you need something that is truly exogenous.

- "No causation without manipulation" (Holland 1986)
- You cannot estimate causal effects of immutable characteristics (gender, race, age, etc.)
- What does the causal effect of gender mean? Problem: CONFOUNDING
- For example, consider whether you can estimate the causal effect of having a discussion leader with certain preferences on deliberation outcomes (Humphreys et al 2006).
- Are networks immutable?

The key is that you need something that is truly exogenous.

- "No causation without manipulation" (Holland 1986)
- You cannot estimate causal effects of immutable characteristics (gender, race, age, etc.)
- What does the causal effect of gender mean? Problem: CONFOUNDING
- For example, consider whether you can estimate the causal effect of having a discussion leader with certain preferences on deliberation outcomes (Humphreys et al 2006).
- Are networks immutable?

The key is that you need something that is truly exogenous.

- "No causation without manipulation" (Holland 1986)
- You cannot estimate causal effects of immutable characteristics (gender, race, age, etc.)
- What does the causal effect of gender mean? Problem: CONFOUNDING
- For example, consider whether you can estimate the causal effect of having a discussion leader with certain preferences on deliberation outcomes (Humphreys et al 2006).
- Are networks immutable?

Field Experimentists Are Concerned about SUTVA

- Randomized field experiments frequently take place in concentrated geographic regions.
- Many subjects in experiments are exposed to multiple other individuals in both the treatment and control groups.
- If individuals are exposed to the treatment indirectly, then estimates of the direct effect of the experiment – the local average treatment effect – may be biased.

Field Experimentists Are Concerned about SUTVA

- Randomized field experiments frequently take place in concentrated geographic regions.
- Many subjects in experiments are exposed to multiple other individuals in both the treatment and control groups.
- If individuals are exposed to the treatment indirectly, then estimates of the direct effect of the experiment – the local average treatment effect – may be biased.

Field Experimentists Are Concerned about SUTVA

- Randomized field experiments frequently take place in concentrated geographic regions.
- Many subjects in experiments are exposed to multiple other individuals in both the treatment and control groups.
- If individuals are exposed to the treatment indirectly, then estimates of the direct effect of the experiment – the local average treatment effect – may be biased.

Bias and Assumptions

Subject to certain restrictions or assumptions it is possible to estimate the direct effect without bias. These include:

- The stable unit treatment value assumption (Holland 1986).
- Designing the experiment to estimate and account for spillover within specified hierarchical groups (Sinclair et al 2012; Hudges and Halloran 2008).
- Limiting the exposure possibilities (Aronow and Samii 2011).

Bias and Assumptions

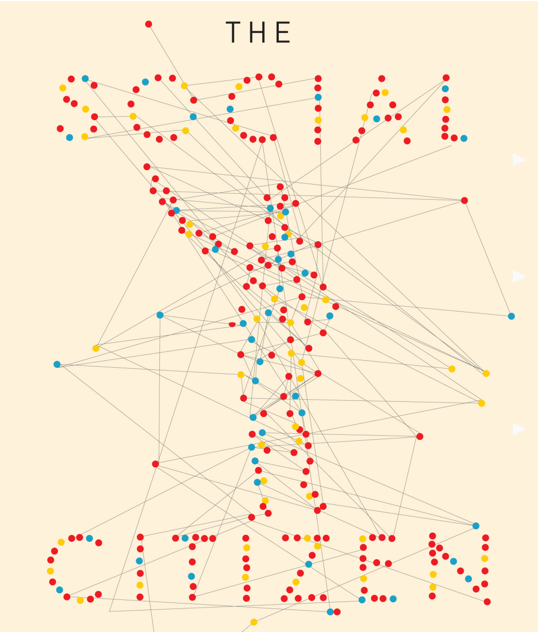
Subject to certain restrictions or assumptions it is possible to estimate the direct effect without bias. These include:

- The stable unit treatment value assumption (Holland 1986).
- Designing the experiment to estimate and account for spillover within specified hierarchical groups (Sinclair et al 2012; Hudges and Halloran 2008).
- Limiting the exposure possibilities (Aronow and Samii 2011).

Bias and Assumptions

Subject to certain restrictions or assumptions it is possible to estimate the direct effect without bias. These include:

- The stable unit treatment value assumption (Holland 1986).
- Designing the experiment to estimate and account for spillover within specified hierarchical groups (Sinclair et al 2012; Hudges and Halloran 2008).
- Limiting the exposure possibilities (Aronow and Samii 2011).



- Experimental evidence on voter turnout (Sinclair 2012).
- Randomly assigning the stimulus in the network (multi-level GOTV campaign).
- Randomly assigning who delivers the stimulus (neighbors vs strangers).

- Experimental evidence on voter turnout (Sinclair 2012)
- Randomly assigning the stimulus in the network (multi-level GOTV campaign).
- Randomly assigning who delivers the stimulus (neighbors vs strangers).

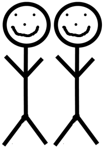
- Experimental evidence on voter turnout (Sinclair 2012)
- Randomly assigning the stimulus in the network (multi-level GOTV campaign).
- Randomly assigning who delivers the stimulus (neighbors vs strangers).

- Experimental evidence on voter turnout (Sinclair 2012)
- Randomly assigning the stimulus in the network (multi-level GOTV campaign).
- Randomly assigning who delivers the stimulus (neighbors vs strangers).

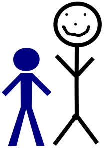
- Experimental evidence on voter turnout (Sinclair 2012)
- Randomly assigning the stimulus in the network (multi-level GOTV campaign).
- Randomly assigning who delivers the stimulus (neighbors vs strangers).

- Experimental evidence on voter turnout (Sinclair 2012)
- Randomly assigning the stimulus in the network (multi-level GOTV campaign).
- Randomly assigning who delivers the stimulus (neighbors vs strangers).

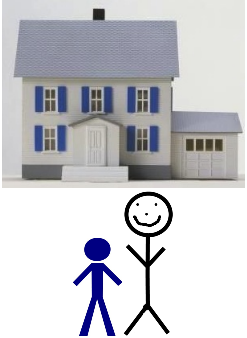
Classic Experiment



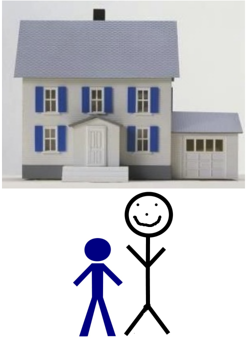
Classic Experiment

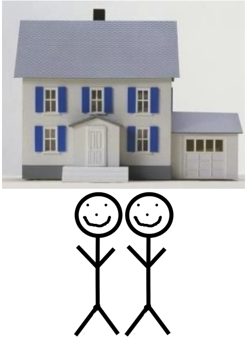


Problem with the Classic Experiment

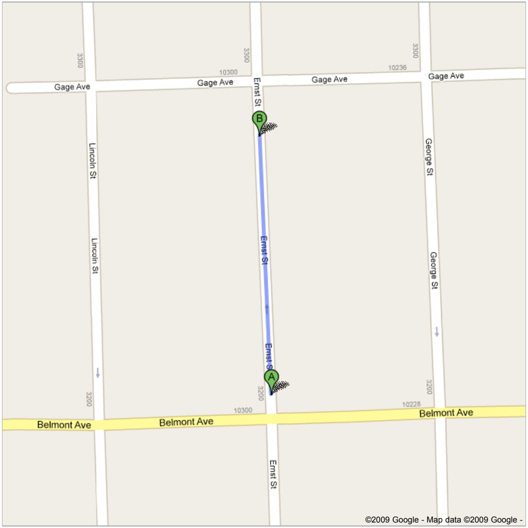


Multilevel Experiment

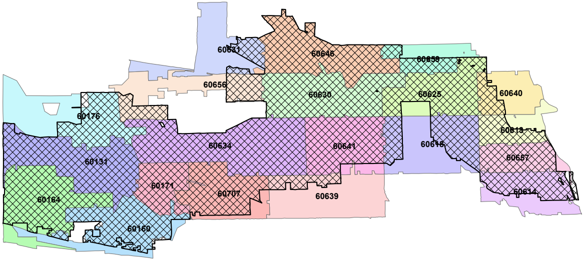




Zip Code 60131-1503

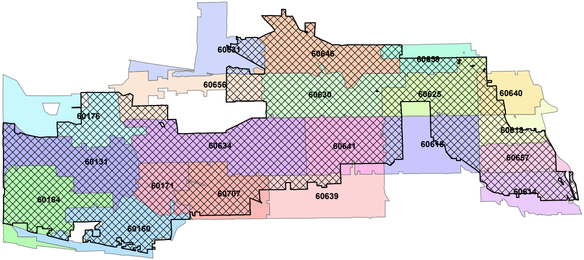


GOTV Multilevel Experimental Design: 5th CD of IL



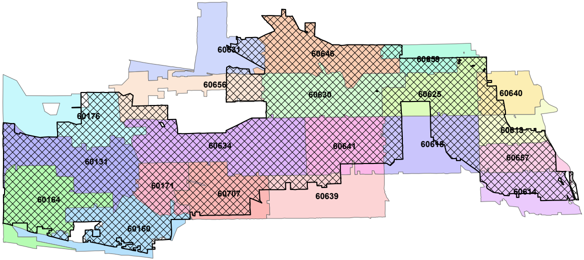
- Mobilization campaign for the 2009 special election.
- Low salience election: one contest and 12% turnout.
- Individuals, households and 9-digit zip codes are randomly assigned to treatment and control.

GOTV Multilevel Experimental Design: 5th CD of IL



- Mobilization campaign for the 2009 special election.
- Low salience election: one contest and 12% turnout.
- Individuals, households and 9-digit zip codes are randomly assigned to treatment and control.

GOTV Multilevel Experimental Design: 5th CD of IL



- Mobilization campaign for the 2009 special election.
- Low salience election: one contest and 12% turnout.
- Individuals, households and 9-digit zip codes are randomly assigned to treatment and control.

Dear Richard L Jensen:

DO YOUR CIVIC DUTY AND VOTE ON APRIL 7!

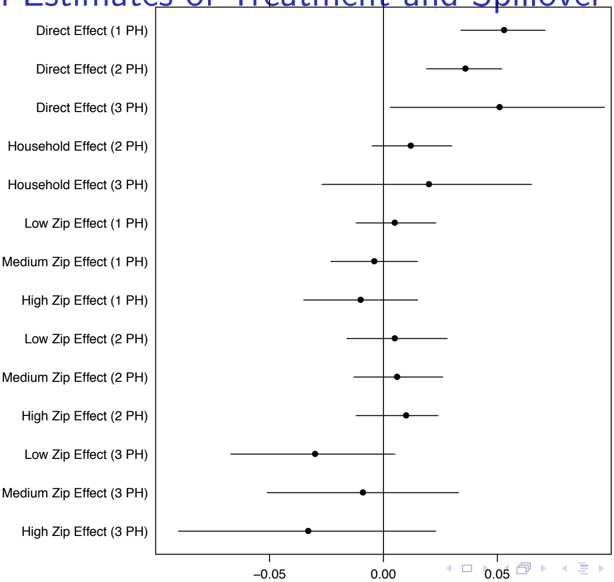
Why do so many people fail to vote? We've been talking about this problem for years, but it only seems to get worse -- especially when elections are held in the spring.

This year we're taking a different approach. We're reminding people that who votes is a matter of public record. The chart shows your name from the list of registered voters and whether you voted in the last two spring elections. The chart also contains an empty space that we will fill in based on whether you vote in the April 7th election.

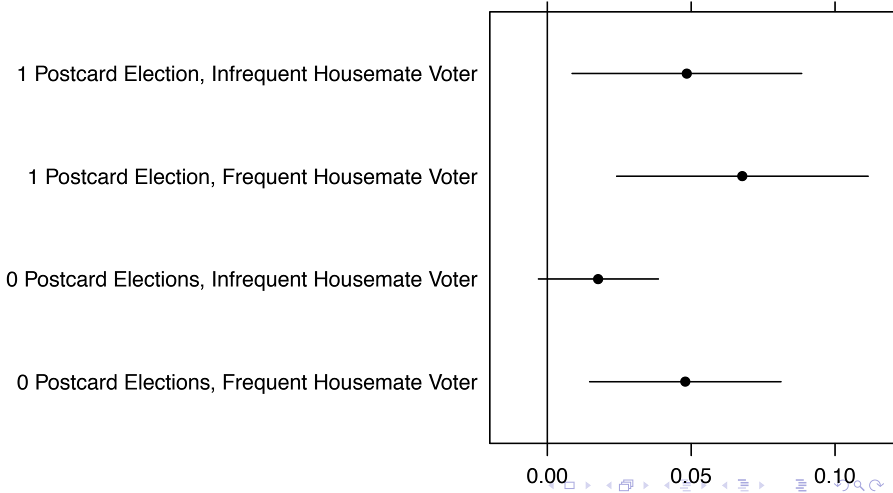
DO YOUR CIVIC DUTY AND VOTE ON APRIL 7!

VOTER NAME	Spring 2006	Spring 2008	April 7
RICHARD L JENSEN	Didn't Vote	Didn't Vote	________

Regression Estimates of Treatment and Spillover Effects

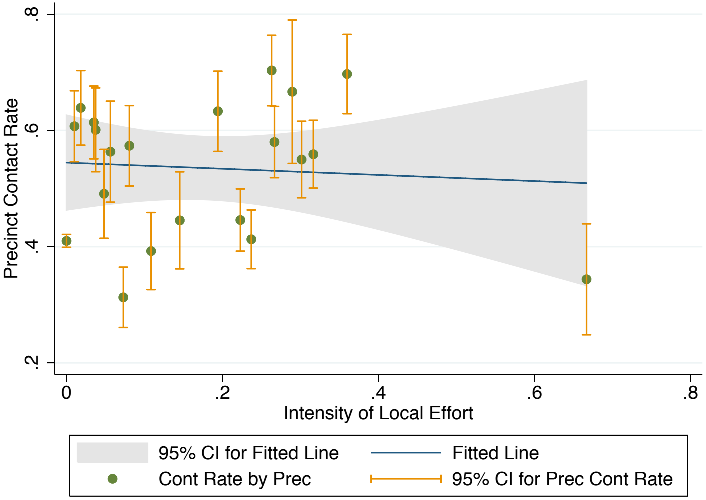


Regression Estimates of Treatment by Housemate Vote History and Postcard Participation

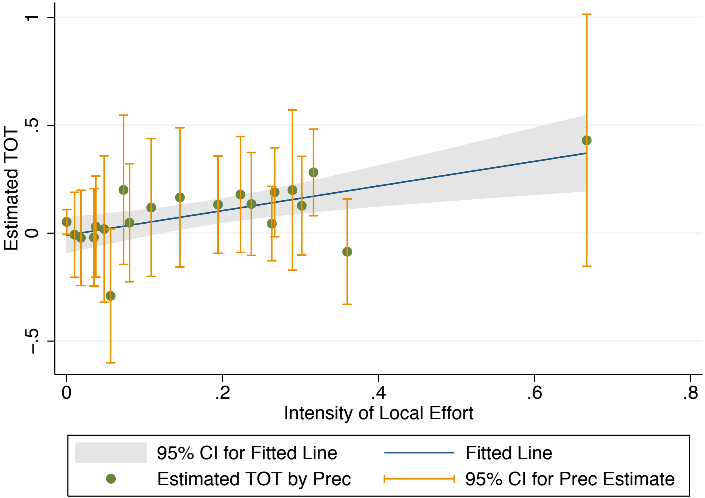


To what extent, if any, is get-out-the-vote face-to-face canvassing by a *neighbor* more effective at increasing turnout than by a *stranger*?

Contact Rate by Percent Local Canvassing Per Precinct



Treatment-on-Treated Effect by Percent Local Canvassing Per Precinct



OLS Coefficients: Effect of Local Contact on Turnout for Contacted Individuals

Variable	OLS	OLS with Control Vars.	OLS with Walker Fixed Effects	OLS with Walker Fixed Effects and Control Vars.
Local Contact	0.056*	0.044**	0.113**	0.094*
	(0.024)	(0.025)	(0.031)	(0.031)
Democratic Registration		0.009		0.011
		(0.015)		(0.015)
Age		0.001*		0.001
		(0.000)		(0.000)
Missing Age		-0.025		-0.022
		(0.036)		(0.036)
Female		-0.015		-0.014
		(0.018)		(0.018)
Missing Female		0.079		0.065
		(0.147)		(0.148)
Vote History		0.175**		0.174**
		(0.007)		(0.007)
Latino		0.127**		0.122**
		(0.023)		(0.023)
Precinct Fixed Effects		Included		Included
Walker Fixed Effects			Included	Included
Constant	0.307**	0.108**	0.500**	0.215
	(0.050)	(0.023)	(0.174)	(0.163)
F(Covariates)	5.23	17.16	3.05	10.28
* $= \alpha = .05$, $= \alpha = .10$				
N	5343	5343	5343	5343

Experimental Findings

- We observe a 5-6 percentage point effect from canvassing.
- We observe a significant marginal effect from home-turf canvassing *above* the baseline effect of canvassing, approximately a 3 percentage point effect.
- Neighbors mobilize more than strangers

Experimental Findings

- We observe a 5-6 percentage point effect from canvassing.
- We observe a significant marginal effect from home-turf canvassing *above* the baseline effect of canvassing, approximately a 3 percentage point effect.
- Neighbors mobilize more than strangers

Experimental Findings

- We observe a 5-6 percentage point effect from canvassing.
- We observe a significant marginal effect from home-turf canvassing *above* the baseline effect of canvassing, approximately a 3 percentage point effect.
- Neighbors mobilize more than strangers

Other Experimental Trials on Voter Turnout Networks

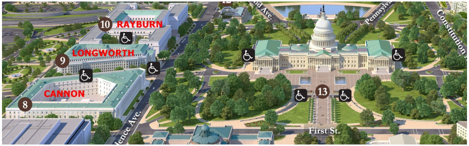
- Nickerson (2008): 60% of the effect spills over to the other household resident.
- Bond et al (2012): close friends approximately 1% more likely to vote as a consequence of a Facebook social network encouragement.

Other Experimental Trials on Voter Turnout Networks

- Nickerson (2008): 60% of the effect spills over to the other household resident.
- Bond et al (2012): close friends approximately 1% more likely to vote as a consequence of a Facebook social network encouragement.

Keep Your Eyes Open For Other Sources of Exogeneity

- House of Representatives Office Buildings: Cannon (5 floors), Longworth (8 floors), Rayburn (4 floors)



Video Evidence: Legislators Have Common Preferences Over Offices

- Data in the video limited to the 103-112 Congresses
- Three buildings: Cannon, Longworth and Rayburn.
- Floors are ordered: C5 is the 5th floor of Cannon.
- Facing the figure, the right-side of Cannon and Longworth have views of the capital.
- Color-coded (ROYGBIV) members by cohort: freshmen are yellow in 103 (and older cohorts are orange). Leaders are black.

Video Evidence: Legislators Have Common Preferences Over Offices

- Data in the video limited to the 103-112 Congresses
- Three buildings: Cannon, Longworth and Rayburn.
- Floors are ordered: C5 is the 5th floor of Cannon.
- Facing the figure, the right-side of Cannon and Longworth have views of the capital.
- Color-coded (ROYGBIV) members by cohort: freshmen are yellow in 103 (and older cohorts are orange). Leaders are black.

Video Evidence: Legislators Have Common Preferences Over Offices

- Data in the video limited to the 103-112 Congresses
- Three buildings: Cannon, Longworth and Rayburn.
- Floors are ordered: C5 is the 5th floor of Cannon.
- Facing the figure, the right-side of Cannon and Longworth have views of the capital.
- Color-coded (ROYGBIV) members by cohort: freshmen are yellow in 103 (and older cohorts are orange). Leaders are black.

Video Evidence: Legislators Have Common Preferences Over Offices

- Data in the video limited to the 103-112 Congresses
- Three buildings: Cannon, Longworth and Rayburn.
- Floors are ordered: C5 is the 5th floor of Cannon.
- Facing the figure, the right-side of Cannon and Longworth have views of the capital.
- Color-coded (ROYGBIV) members by cohort: freshmen are yellow in 103 (and older cohorts are orange). Leaders are black.

Video Evidence: Legislators Have Common Preferences Over Offices

- Data in the video limited to the 103-112 Congresses
- Three buildings: Cannon, Longworth and Rayburn.
- Floors are ordered: C5 is the 5th floor of Cannon.
- Facing the figure, the right-side of Cannon and Longworth have views of the capital.
- Color-coded (ROYGBIV) members by cohort: freshmen are yellow in 103 (and older cohorts are orange). Leaders are black.

Common Preferences Over Offices

Video

Outline

Questions for Research

- Do you need to see the whole network in order to understand network effects?
- Can networks be studied in a more controlled laboratory environment to better understand their structural implications? (a la Nick Weller)
- What is the appropriate role for formal models of social and economic networks in this research?

Questions for Research

- Do you need to see the whole network in order to understand network effects?
- Can networks be studied in a more controlled laboratory environment to better understand their structural implications? (a la Nick Weller)
- What is the appropriate role for formal models of social and economic networks in this research?

Questions for Research

- Do you need to see the whole network in order to understand network effects?
- Can networks be studied in a more controlled laboratory environment to better understand their structural implications? (a la Nick Weller)
- What is the appropriate role for formal models of social and economic networks in this research?

Citizens are part of their social environment.

- We typically characterize citizens as atomistic actors, yet the most basic action in a participatory democracy – the decision to turn out to cast a ballot – is a decision associated with that citizen's social network.
- Social networks lead to particular political democratic outcomes.

Citizens are part of their social environment.

- We typically characterize citizens as atomistic actors, yet the most basic action in a participatory democracy – the decision to turn out to cast a ballot – is a decision associated with that citizen's social network.
- Social networks lead to particular political democratic outcomes.

Why should social scientists study networks?

We are all connected.

- Deepens our understanding of individual behavior and attitudes.
- Helps us formalize the relationships across political institutions.
- Provides another tool to observe change in political institutions (such as donors to elected representatives).

Why should social scientists study networks?

We are all connected.

- Deepens our understanding of individual behavior and attitudes.
- Helps us formalize the relationships across political institutions.
- Provides another tool to observe change in political institutions (such as donors to elected representatives).

Why should social scientists study networks?

We are all connected.

- Deepens our understanding of individual behavior and attitudes.
- Helps us formalize the relationships across political institutions.
- Provides another tool to observe change in political institutions (such as donors to elected representatives).

Why should social scientists study networks?

We are all connected.

- Deepens our understanding of individual behavior and attitudes.
- Helps us formalize the relationships across political institutions.
- Provides another tool to observe change in political institutions (such as donors to elected representatives).

Super Influencers

- Studying networks isn't only about academic research. It's also about understanding how your political voice matters, and what you can do to be heard.

Practical Matters: Available Data

- Mark Newman's Network Data
- Stanford Large Network Dataset Collection

Practical Matters: Available Data

- Mark Newman's Network Data
- Stanford Large Network Dataset Collection

Introduction

- Randomized field experiments frequently take place in concentrated geographic regions.
- Many subjects in experiments are exposed to multiple other individuals in both the treatment and control groups.
- If individuals are exposed to the treatment indirectly, then estimates of the direct effect of the experiment – the local average treatment effect – may be biased.
- What techniques can we develop to:
    - estimate indirect effects and
    - control for their influence so as to be able to estimate the direct effect without bias?

Introduction

- Randomized field experiments frequently take place in concentrated geographic regions.
- Many subjects in experiments are exposed to multiple other individuals in both the treatment and control groups.
- If individuals are exposed to the treatment indirectly, then estimates of the direct effect of the experiment – the local average treatment effect – may be biased.
- What techniques can we develop to:
  - estimate indirect effects and
  - control for their influence so as to be able to estimate the direct effect without bias?

Introduction

- Randomized field experiments frequently take place in concentrated geographic regions.
- Many subjects in experiments are exposed to multiple other individuals in both the treatment and control groups.
- If individuals are exposed to the treatment indirectly, then estimates of the direct effect of the experiment – the local average treatment effect – may be biased.
- What techniques can we develop to:
  - estimate indirect effects and
  - control for their influence so as to be able to estimate the direct effect without bias?

Introduction

- Randomized field experiments frequently take place in concentrated geographic regions.
- Many subjects in experiments are exposed to multiple other individuals in both the treatment and control groups.
- If individuals are exposed to the treatment indirectly, then estimates of the direct effect of the experiment – the local average treatment effect – may be biased.
- What techniques can we develop to:
    - estimate indirect effects and
    - control for their influence so as to be able to estimate the direct effect without bias?

Introduction

- Randomized field experiments frequently take place in concentrated geographic regions.
- Many subjects in experiments are exposed to multiple other individuals in both the treatment and control groups.
- If individuals are exposed to the treatment indirectly, then estimates of the direct effect of the experiment – the local average treatment effect – may be biased.
- What techniques can we develop to:
  - estimate indirect effects and
  - control for their influence so as to be able to estimate the direct effect without bias?

Introduction

- Randomized field experiments frequently take place in concentrated geographic regions.
- Many subjects in experiments are exposed to multiple other individuals in both the treatment and control groups.
- If individuals are exposed to the treatment indirectly, then estimates of the direct effect of the experiment – the local average treatment effect – may be biased.
- What techniques can we develop to:
  - estimate indirect effects and
  - control for their influence so as to be able to estimate the direct effect without bias?

Bias, Assumptions, and Statement of the Problem

- Subject to certain restrictions or assumptions it is possible to estimate the direct effect without bias. These include:
  - Making the stable unit treatment value assumption (Holland 1986).
  - Designing the experiment to estimate and account for spillover within specified hierarchical groups (Sinclair et al 2012; Hudges and Halloran 2008).
  - Limiting the exposure possibilities (Aronow and Samii 2011).
  - Explicitly testing a specific model of exposure (Bowers, Fredrickson and Panagopoulos 2012).
- Yet, it is not clear how to estimate treatment effects if there is heterogeneity in the structure of the network that is correlated with heterogeneity in susceptibility to indirect treatment.

Bias, Assumptions, and Statement of the Problem

- Subject to certain restrictions or assumptions it is possible to estimate the direct effect without bias. These include:
  - Making the stable unit treatment value assumption (Holland 1986).
  - Designing the experiment to estimate and account for spillover within specified hierarchical groups (Sinclair et al 2012; Hudges and Halloran 2008).
  - Limiting the exposure possibilities (Aronow and Samii 2011).
  - Explicitly testing a specific model of exposure (Bowers, Fredrickson and Panagopoulos 2012).
- Yet, it is not clear how to estimate treatment effects if there is heterogeneity in the structure of the network that is correlated with heterogeneity in susceptibility to indirect treatment.

Bias, Assumptions, and Statement of the Problem

- Subject to certain restrictions or assumptions it is possible to estimate the direct effect without bias. These include:
  - Making the stable unit treatment value assumption (Holland 1986).
  - Designing the experiment to estimate and account for spillover within specified hierarchical groups (Sinclair et al 2012; Hudges and Halloran 2008).
  - Limiting the exposure possibilities (Aronow and Samii 2011).
  - Explicitly testing a specific model of exposure (Bowers, Fredrickson and Panagopoulos 2012).
- Yet, it is not clear how to estimate treatment effects if there is heterogeneity in the structure of the network that is correlated with heterogeneity in susceptibility to indirect treatment.

Bias, Assumptions, and Statement of the Problem

- Subject to certain restrictions or assumptions it is possible to estimate the direct effect without bias. These include:
  - Making the stable unit treatment value assumption (Holland 1986).
  - Designing the experiment to estimate and account for spillover within specified hierarchical groups (Sinclair et al 2012; Hudges and Halloran 2008).
  - Limiting the exposure possibilities (Aronow and Samii 2011).
  - Explicitly testing a specific model of exposure (Bowers, Fredrickson and Panagopoulos 2012).
- Yet, it is not clear how to estimate treatment effects if there is heterogeneity in the structure of the network that is correlated with heterogeneity in susceptibility to indirect treatment.

Bias, Assumptions, and Statement of the Problem

- Subject to certain restrictions or assumptions it is possible to estimate the direct effect without bias. These include:
  - Making the stable unit treatment value assumption (Holland 1986).
  - Designing the experiment to estimate and account for spillover within specified hierarchical groups (Sinclair et al 2012; Hudges and Halloran 2008).
  - Limiting the exposure possibilities (Aronow and Samii 2011).
  - Explicitly testing a specific model of exposure (Bowers, Fredrickson and Panagopoulos 2012).
- Yet, it is not clear how to estimate treatment effects if there is heterogeneity in the structure of the network that is correlated with heterogeneity in susceptibility to indirect treatment.

Bias, Assumptions, and Statement of the Problem

- Subject to certain restrictions or assumptions it is possible to estimate the direct effect without bias. These include:
  - Making the stable unit treatment value assumption (Holland 1986).
  - Designing the experiment to estimate and account for spillover within specified hierarchical groups (Sinclair et al 2012; Hudges and Halloran 2008).
  - Limiting the exposure possibilities (Aronow and Samii 2011).
  - Explicitly testing a specific model of exposure (Bowers, Fredrickson and Panagopoulos 2012).
- Yet, it is not clear how to estimate treatment effects if there is heterogeneity in the structure of the network that is correlated with heterogeneity in susceptibility to indirect treatment.

This Project

- Establishes an empirical technique which leverages the community detection process *modularity maximization* to estimate indirect and direct treatment effects in a randomized field experiment where treatment is randomly assigned to nodes in a fixed network.
- Estimates two types of indirect effects: within-community effects and across-community effects

This Project

- Establishes an empirical technique which leverages the community detection process *modularity maximization* to estimate indirect and direct treatment effects in a randomized field experiment where treatment is randomly assigned to nodes in a fixed network.
- Estimates two types of indirect effects: within-community effects and across-community effects

Community Detection

- In the field of network analysis there are many possible algorithms to detect *communities* within observed networks.
- Represent a set of individuals as nodes of a graph and their relationships as vertices.
- A *network community* is a subset of a social network graph that is more internally connected than externally connected.

Community Detection

- In the field of network analysis there are many possible algorithms to detect *communities* within observed networks.
- Represent a set of individuals as nodes of a graph and their relationships as vertices.
- A *network community* is a subset of a social network graph that is more internally connected than externally connected.

Community Detection

- In the field of network analysis there are many possible algorithms to detect *communities* within observed networks.
- Represent a set of individuals as nodes of a graph and their relationships as vertices.
- A *network community* is a subset of a social network graph that is more internally connected than externally connected.

Community Detection

- Community detection algorithms do not presuppose knowledge of the number and size of the underlying groups.
- Find non-overlapping communities.
- Rely upon *modularity maximization*, formalized by Girvan and Newman (2002).

Community Detection

- Community detection algorithms do not presuppose knowledge of the number and size of the underlying groups.
- Find non-overlapping communities.
- Rely upon *modularity maximization*, formalized by Girvan and Newman (2002).

Community Detection

- Community detection algorithms do not presuppose knowledge of the number and size of the underlying groups.
- Find non-overlapping communities.
- Rely upon *modularity maximization*, formalized by Girvan and Newman (2002).

Modularity

Let $\mathbf{e}$ be a symmetric $n \times n$ matrix where each element e_{ij} represents the fraction of all edges that link community i to vertices in community j. Modularity is defined, then, as:

$$Q = \sum_i (e_{ii} - a_i^2) = \text{Tr } \mathbf{e} - ||\mathbf{e}^2||$$

where a_i is defined as the row sums, $a_i = \sum_j e_{ij}$. Tr $\mathbf{e}$, the trace of $\mathbf{e}$ gives the sum of edges connecting vertices in the same community, and $||\mathbf{e}^2||$ represents the sum of the elements of the matrix $\mathbf{e}^2$.

- We can think of $||\mathbf{e}^2||$ as measuring, holding the detected community structure fixed, the expected number of edges connecting communities if connections between vertices were random.
- Modularity, Q, can then take on values between 0 and 1.

Modularity

Let $\mathbf{e}$ be a symmetric $n \times n$ matrix where each element e_{ij} represents the fraction of all edges that link community i to vertices in community j. Modularity is defined, then, as:

$$Q = \sum_i (e_{ii} - a_i^2) = \text{Tr } \mathbf{e} - ||\mathbf{e}^2||$$

where a_i is defined as the row sums, $a_i = \sum_j e_{ij}$. Tr $\mathbf{e}$, the trace of $\mathbf{e}$ gives the sum of edges connecting vertices in the same community, and $||\mathbf{e}^2||$ represents the sum of the elements of the matrix $\mathbf{e}^2$.

- We can think of $||\mathbf{e}^2||$ as measuring, holding the detected community structure fixed, the expected number of edges connecting communities if connections between vertices were random.
- Modularity, Q, can then take on values between 0 and 1.

Modularity

Let $\mathbf{e}$ be a symmetric $n \times n$ matrix where each element e_{ij} represents the fraction of all edges that link community i to vertices in community j. Modularity is defined, then, as:

$$Q = \sum_i (e_{ii} - a_i^2) = \text{Tr } \mathbf{e} - ||\mathbf{e}^2||$$

where a_i is defined as the row sums, $a_i = \sum_j e_{ij}$. Tr $\mathbf{e}$, the trace of $\mathbf{e}$ gives the sum of edges connecting vertices in the same community, and $||\mathbf{e}^2||$ represents the sum of the elements of the matrix $\mathbf{e}^2$.

- We can think of $||\mathbf{e}^2||$ as measuring, holding the detected community structure fixed, the expected number of edges connecting communities if connections between vertices were random.
- Modularity, Q, can then take on values between 0 and 1.

Steps of Modularity Maximization

1. Calculate betweenness (the number of shortest paths that run along a given edge) for every edge.
2. The edge with the highest betweenness score is removed, as this edge is most likely to connect communities as opposed to lie within communities.
3. Estimate modularity (measures the number of within-community edges relative to a null model of a random graph with the same degree distribution).
4. The betweenness scores are recalculated within the new graph, and the process repeats.
5. Continue to remove edges until modularity is maximized.

Steps of Modularity Maximization

1. Calculate betweenness (the number of shortest paths that run along a given edge) for every edge.
2. The edge with the highest betweenness score is removed, as this edge is most likely to connect communities as opposed to lie within communities.
3. Estimate modularity (measures the number of within-community edges relative to a null model of a random graph with the same degree distribution).
4. The betweenness scores are recalculated within the new graph, and the process repeats.
5. Continue to remove edges until modularity is maximized.

Steps of Modularity Maximization

1. Calculate betweenness (the number of shortest paths that run along a given edge) for every edge.
2. The edge with the highest betweenness score is removed, as this edge is most likely to connect communities as opposed to lie within communities.
3. Estimate modularity (measures the number of within-community edges relative to a null model of a random graph with the same degree distribution).
4. The betweenness scores are recalculated within the new graph, and the process repeats.
5. Continue to remove edges until modularity is maximized.

Steps of Modularity Maximization

1. Calculate betweenness (the number of shortest paths that run along a given edge) for every edge.
2. The edge with the highest betweenness score is removed, as this edge is most likely to connect communities as opposed to lie within communities.
3. Estimate modularity (measures the number of within-community edges relative to a null model of a random graph with the same degree distribution).
4. The betweenness scores are recalculated within the new graph, and the process repeats.
5. Continue to remove edges until modularity is maximized.

Steps of Modularity Maximization

1. Calculate betweenness (the number of shortest paths that run along a given edge) for every edge.
2. The edge with the highest betweenness score is removed, as this edge is most likely to connect communities as opposed to lie within communities.
3. Estimate modularity (measures the number of within-community edges relative to a null model of a random graph with the same degree distribution).
4. The betweenness scores are recalculated within the new graph, and the process repeats.
5. Continue to remove edges until modularity is maximized.

Model

- We have finite n individuals i, where $i = \{1, 2, \ldots, n\}$ who reside in a political network G that is a strongly-connected, undirected graph.
- Edges in G represent shared political choices between individuals. Link intensity increases as the number of shared choices increases.
- The modularity maximization network detection algorithm generates a length n vector N of community membership, where each individual falls into a single community.

Model

- We have finite n individuals i, where $i = \{1, 2, \ldots, n\}$ who reside in a political network G that is a strongly-connected, undirected graph.
- Edges in G represent shared political choices between individuals. Link intensity increases as the number of shared choices increases.
- The modularity maximization network detection algorithm generates a length n vector N of community membership, where each individual falls into a single community.

Model

- We have finite n individuals i, where $i = \{1, 2, \ldots, n\}$ who reside in a political network G that is a strongly-connected, undirected graph.
- Edges in G represent shared political choices between individuals. Link intensity increases as the number of shared choices increases.
- The modularity maximization network detection algorithm generates a length n vector N of community membership, where each individual falls into a single community.

Model

Using G and N, we write out our $n \times n$ social transition matrix M such that each individual entry is weighted by whether individuals i and j are members in the same community or different communities. That is, we define a function $c(N_i, N_j) s.t. \forall i, j \in \{1, \ldots, n\}$

$$c(N_i, N_j) = \begin{cases} \alpha & \text{if } N_i = N_j \text{ for } i \neq j \\ \beta & \text{if } N_i \neq N_j \text{ for } i \neq j \\ 1 & \text{for } i = j \end{cases}$$

We can then define our social transition matrix

$$M_{n,n} = \begin{pmatrix} c(N_1, N_1) g_{1,1} & c(N_1, N_2) g_{1,2} & \cdots & c(N_1, N_n) g_{1,n} \\ c(N_2, N_1) g_{2,1} & c(N_2, N_2) g_{2,2} & \cdots & c(N_2, N_n) g_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ c(N_n, N_1) g_{n,1} & c(N_n, N_2) g_{n,2} & \cdots & c(N_n, N_n) g_{n,n} \end{pmatrix}$$

Estimation

Individuals are randomly assigned a binary treatment T for each $i \in \{1, \ldots, n\}$. We can write this assignment as a vector of length n called $d(i)$ where

$$d(i) = \begin{cases} \delta & \text{if } T_i = 1 \\ 0 & \text{if } T_i = 0 \end{cases}$$

We want to know how this external stimulus percolates throughout the social network, and in particular if there is evidence of network influence in an individual's political decisions as a consequence of this stimulus.

DeGroot Updating: Turnout Example

- We know the turnout decisions for all n voters in the election, and we represent these choices in a vector v.
- If social networks are influential for an individual's decision, and if individuals update based upon a weighted average of the turnout probabilities of their social network neighbors, then given that we know who actually cast a ballot in the election we are able to write an equation for this process.
- That is, $Pr(v_i) = m_{i,1}d(1) + m_{i,2}d(2) + \ldots + m_{i,n}d(n) + \epsilon_i \forall i \in \{1, \ldots, n\}$.

DeGroot Updating: Turnout Example

- We know the turnout decisions for all n voters in the election, and we represent these choices in a vector v.
- If social networks are influential for an individual's decision, and if individuals update based upon a weighted average of the turnout probabilities of their social network neighbors, then given that we know who actually cast a ballot in the election we are able to write an equation for this process.
- That is, $Pr(v_i) = m_{i,1}d(1) + m_{i,2}d(2) + \ldots + m_{i,n}d(n) + \epsilon_i \forall i \in \{1, \ldots, n\}$.

DeGroot Updating: Turnout Example

- We know the turnout decisions for all n voters in the election, and we represent these choices in a vector v.
- If social networks are influential for an individual's decision, and if individuals update based upon a weighted average of the turnout probabilities of their social network neighbors, then given that we know who actually cast a ballot in the election we are able to write an equation for this process.
- That is, $Pr(v_i) = m_{i,1}d(1) + m_{i,2}d(2) + \ldots + m_{i,n}d(n) + \epsilon_i \forall i \in \{1, \ldots, n\}$.

Estimation

We estimate three parameters:

- the *direct effect* (δ) of receiving the treatment,
- the *average within-community indirect effect* (α) from one individual who has received the treatment to another who has not and
- the *average across-community indirect effect* (β) from one individual who has received the treatment to another who has not and is a member of a different community.

Estimation

We estimate three parameters:

- the *direct effect* (δ) of receiving the treatment,
- the *average within-community indirect effect* (α) from one individual who has received the treatment to another who has not and
- the *average across-community indirect effect* (β) from one individual who has received the treatment to another who has not and is a member of a different community.

Estimation

We estimate three parameters:

- the *direct effect* (δ) of receiving the treatment,
- the *average within-community indirect effect* (α) from one individual who has received the treatment to another who has not and
- the *average across-community indirect effect* (β) from one individual who has received the treatment to another who has not and is a member of a different community.

Estimation

To review, we follow the steps below:

1. Calculate a social graph G given a set of shared political behaviors
2. Apply the *modularity maximization* algorithm to detect the communities vector N
3. Use G and N to calculate the social transition matrix M
4. Set $M \cdot d$ equal to our observed outcome v and estimate our parameters.

Estimation

To review, we follow the steps below:

1. Calculate a social graph G given a set of shared political behaviors
2. Apply the *modularity maximization* algorithm to detect the communities vector N
3. Use G and N to calculate the social transition matrix M
4. Set $M \cdot d$ equal to our observed outcome v and estimate our parameters.

Estimation

To review, we follow the steps below:

1. Calculate a social graph G given a set of shared political behaviors
2. Apply the *modularity maximization* algorithm to detect the communities vector N
3. Use G and N to calculate the social transition matrix M
4. Set $M \cdot d$ equal to our observed outcome v and estimate our parameters.

Estimation

To review, we follow the steps below:

1. Calculate a social graph G given a set of shared political behaviors
2. Apply the *modularity maximization* algorithm to detect the communities vector N
3. Use G and N to calculate the social transition matrix M
4. Set $M \cdot d$ equal to our observed outcome v and estimate our parameters.

Causal Inferences

- One concern with this style of inference is that while individuals are randomly assigned to treatment and control, they are not randomly assigned to their indirect exposures (as they choose whether or not they have friends at all and furthermore whether those friends are members of the same or different communities).
- To account for this we'll have to make an ignorability assumption

Causal Inferences

- One concern with this style of inference is that while individuals are randomly assigned to treatment and control, they are not randomly assigned to their indirect exposures (as they choose whether or not they have friends at all and furthermore whether those friends are members of the same or different communities).
- To account for this we'll have to make an ignorability assumption

Simulation: Network

- Start with 3K nodes assigned to 3 types (40%, 40% and 20% assigned to each)
- Probability that any two nodes of the same type are connected = 0.3
- Probability that any two nodes of different types are connected = 0.1

Simulation: Network

- Start with 3K nodes assigned to 3 types (40%, 40% and 20% assigned to each)
- Probability that any two nodes of the same type are connected = 0.3
- Probability that any two nodes of different types are connected = 0.1

Simulation: Network

- Start with 3K nodes assigned to 3 types (40%, 40% and 20% assigned to each)
- Probability that any two nodes of the same type are connected $= 0.3$
- Probability that any two nodes of different types are connected $= 0.1$

Simulation: Outcome

- Probability of outcome for each individual drawn from a N(0,1) distribution
- Percent of nodes that are randomly assigned to treatment = 20%
- Direct treatment Effect = 0.03
- Within-Community (individual) Treatment Effect = 0.20
- Across-Community (individual) Treatment Effect = 0.01

Simulation: Outcome

- Probability of outcome for each individual drawn from a N(0,1) distribution
- Percent of nodes that are randomly assigned to treatment = 20%
- Direct treatment Effect = 0.03
- Within-Community (individual) Treatment Effect = 0.20
- Across-Community (individual) Treatment Effect = 0.01

Simulation: Outcome

- Probability of outcome for each individual drawn from a N(0,1) distribution
- Percent of nodes that are randomly assigned to treatment = 20%
- Direct treatment Effect = 0.03
- Within-Community (individual) Treatment Effect = 0.20
- Across-Community (individual) Treatment Effect = 0.01

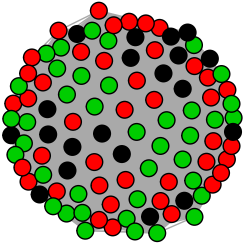
Simulation: Outcome

- Probability of outcome for each individual drawn from a N(0,1) distribution
- Percent of nodes that are randomly assigned to treatment = 20%
- Direct treatment Effect = 0.03
- Within-Community (individual) Treatment Effect = 0.20
- Across-Community (individual) Treatment Effect = 0.01

Simulation: Outcome

- Probability of outcome for each individual drawn from a N(0,1) distribution
- Percent of nodes that are randomly assigned to treatment = 20%
- Direct treatment Effect = 0.03
- Within-Community (individual) Treatment Effect = 0.20
- Across-Community (individual) Treatment Effect = 0.01

Simulation: Community Detection



Simulation: Direct Effects and Average Total Indirect Effects

GLM Regression Coefficients:

	Estimate	Simulated Value
intercept	0.4691*	–
	(0.0858)	
alpha (within community indirect)	0.0021*	.006
	(0.0006)	
beta (across community indirect)	0.0008	.0003
	(0.0013)	
delta (direct)	0.0484*	.03
	(0.0218)	

Sample Mailer

```
Dear Richard L Jensen:

DO YOUR CIVIC DUTY AND VOTE ON APRIL 7!

Why do so many people fail to vote?  We've been talking about
this problem for years, but it only seems to get worse --
especially when elections are held in the spring.

This year we're taking a different approach.  We're reminding
people that who votes is a matter of public record.  The chart
shows your name from the list of registered voters and whether
you voted in the last two spring elections.  The chart also
contains an empty space that we will fill in based on whether
you vote in the April 7th election.

DO YOUR CIVIC DUTY AND VOTE ON APRIL 7!

VOTER NAME                    Spring 2006  Spring 2008   April 7
RICHARD L JENSEN              Didn't Vote  Didn't Vote   _______
```

GLM Regression Coefficients: Direct and Indirect Effects

Variable	Coefficients	Coefficients
Direct Effect	0.037	0.049*
	(.02)	(.02)
Within-Community Indirect Effect	-0.01	-0.0003
	(.0005)	(.001)
Across-Community Indirect Effect	-0.009	-0.003*
	(.001)	(.0008)
Constant	3.33	0.67
	(.189)	(3.70)
Covariates	No	Yes

Covariates include registration year, gender, race, Democratic primary participation, age, household size and vote history.

Summary

- The existing literature is not clear how to estimate treatment effects if there is heterogeneity in the structure of the network that is correlated with heterogeneity in susceptibility to treatment.
- This project establishes an empirical technique which leverages the community detection process *modularity maximization* to estimate indirect and direct treatment effects in a randomized field experiment where treatment is randomly assigned to nodes in a fixed network.
- Estimates two types of indirect effects: within-community effects and across-community effects

Summary

- The existing literature is not clear how to estimate treatment effects if there is heterogeneity in the structure of the network that is correlated with heterogeneity in susceptibility to treatment.
- This project establishes an empirical technique which leverages the community detection process *modularity maximization* to estimate indirect and direct treatment effects in a randomized field experiment where treatment is randomly assigned to nodes in a fixed network.
- Estimates two types of indirect effects: within-community effects and across-community effects

Summary

- The existing literature is not clear how to estimate treatment effects if there is heterogeneity in the structure of the network that is correlated with heterogeneity in susceptibility to treatment.
- This project establishes an empirical technique which leverages the community detection process *modularity maximization* to estimate indirect and direct treatment effects in a randomized field experiment where treatment is randomly assigned to nodes in a fixed network.
- Estimates two types of indirect effects: within-community effects and across-community effects

Extensions

- Raising M to higher powers allows indirect weight to be placed on more distant network connections
- Parameterize $c(N_i, N_j)$ based upon additional theoretical insights (some particular communities should have their own parameters, for example?)
- Finding other empirical examples
- Apply the Newman (2004) modularity algorithm
- Consider a model where indirect exposures are not linearly-additive

Extensions

- Raising M to higher powers allows indirect weight to be placed on more distant network connections
- Parameterize $c(N_i, N_j)$ based upon additional theoretical insights (some particular communities should have their own parameters, for example?)
- Finding other empirical examples
- Apply the Newman (2004) modularity algorithm
- Consider a model where indirect exposures are not linearly-additive

Extensions

- Raising M to higher powers allows indirect weight to be placed on more distant network connections
- Parameterize $c(N_i, N_j)$ based upon additional theoretical insights (some particular communities should have their own parameters, for example?)
- Finding other empirical examples
- Apply the Newman (2004) modularity algorithm
- Consider a model where indirect exposures are not linearly-additive

Extensions

- Raising M to higher powers allows indirect weight to be placed on more distant network connections
- Parameterize $c(N_i, N_j)$ based upon additional theoretical insights (some particular communities should have their own parameters, for example?)
- Finding other empirical examples
- Apply the Newman (2004) modularity algorithm
- Consider a model where indirect exposures are not linearly-additive

Extensions

- Raising M to higher powers allows indirect weight to be placed on more distant network connections
- Parameterize $c(N_i, N_j)$ based upon additional theoretical insights (some particular communities should have their own parameters, for example?)
- Finding other empirical examples
- Apply the Newman (2004) modularity algorithm
- Consider a model where indirect exposures are not linearly-additive

Future Research Design

- We would suggest blocking by the number of within-community membership ties and across-community membership ties instead of randomly assigning individuals to treatment and control when designing experiments in a social network context

Research Agenda

- Part of a larger interest in the role of an individual's social environment on their political behavior
- Important to understand how people influence us who are similar, and how people influence us who are different, and in particular to understand something about how social structure enables different kinds of conversations and kinds of political engagement
- This particular estimation strategy draws our attention to a particular kind of spillover and lets us re-use experimental data to answer another political question.

Research Agenda

- Part of a larger interest in the role of an individual's social environment on their political behavior
- Important to understand how people influence us who are similar, and how people influence us who are different, and in particular to understand something about how social structure enables different kinds of conversations and kinds of political engagement
- This particular estimation strategy draws our attention to a particular kind of spillover and lets us re-use experimental data to answer another political question.

Research Agenda

- Part of a larger interest in the role of an individual's social environment on their political behavior
- Important to understand how people influence us who are similar, and how people influence us who are different, and in particular to understand something about how social structure enables different kinds of conversations and kinds of political engagement
- This particular estimation strategy draws our attention to a particular kind of spillover and lets us re-use experimental data to answer another political question.