POIR 613: Computational Social Science

Pablo Barberá

University of Southern California
pablobarbera.com

Course website:
pablobarbera.com/POIR613/

Social network analysis:
diffusion dynamics

Diffusion dynamics

Diffusion via social ties are key mechanisms explaining how **diseases, information, and behavior spreads**.

Diffusion dynamics

Two types of diffusion processes:

1. Simple contagion
   - One contact is enough for contagion (adopting behavior, receiving information, etc)
   - Example: spread of diseases
2. Complex contagion
   - Multiple and/or diverse contacts are necessary for contagion
   - *Threshold* models: adopt behavior if x% of your ties have already adopted it
   - Examples: online memes, technology or social media adoption, collective action, public opinion change, etc.
   - Most common mechanism in social processes

Contagion dynamics

Why does it matter? Interaction between network properties and diffusion dynamics:

- In highly clustered networks, complex contagion is unlikely to reach the entire network
- Simple contagion will be faster if it reaches a node with degree centrality
- In contrast, individuals with high betweenness centrality are key if contagion is complex

How to study contagion dynamics

Computational models (aka agent-based models) have greatly advanced study of diffusion:

- Models that make assumptions about boundedly ration individual actors (e.g. preferences, behavior, connections)
- Then run simulations to understand how variation in key parameters lead to different global patterns
- A few well-known examples:
  - Schelling's segregation model
  - Axelrod's model of ethnocentrism
- Common insight from agent-based models: asymmetry between individual preferences and collective outcomes
- Examples using Netlogo (see Segregation and Virus)

Social network analysis:
tie strength

Tie strength

Not all ties are created equal:

- Strong ties: family, partner, close friends...
- Weak ties: distant relative, acquaintances, co-workers...

Where tie **strength** can be defined in terms of:

- Frequency of interaction
- Potential to persuade, trust
- Shared traits
- Many mutual contacts

The strength of weak ties

Granovetter (1973, AJS):

- Random sample of recent job changers in Boston
- "How often did you see the contact around the time they passed job information?" (measure of tie strength)
- Key finding: 55.6% saw contact only occasionally
- The strength of weak ties – Why?
  1. Less influential, but strength in numbers
  2. Bridges across loosely connected network components

The strength of weak ties

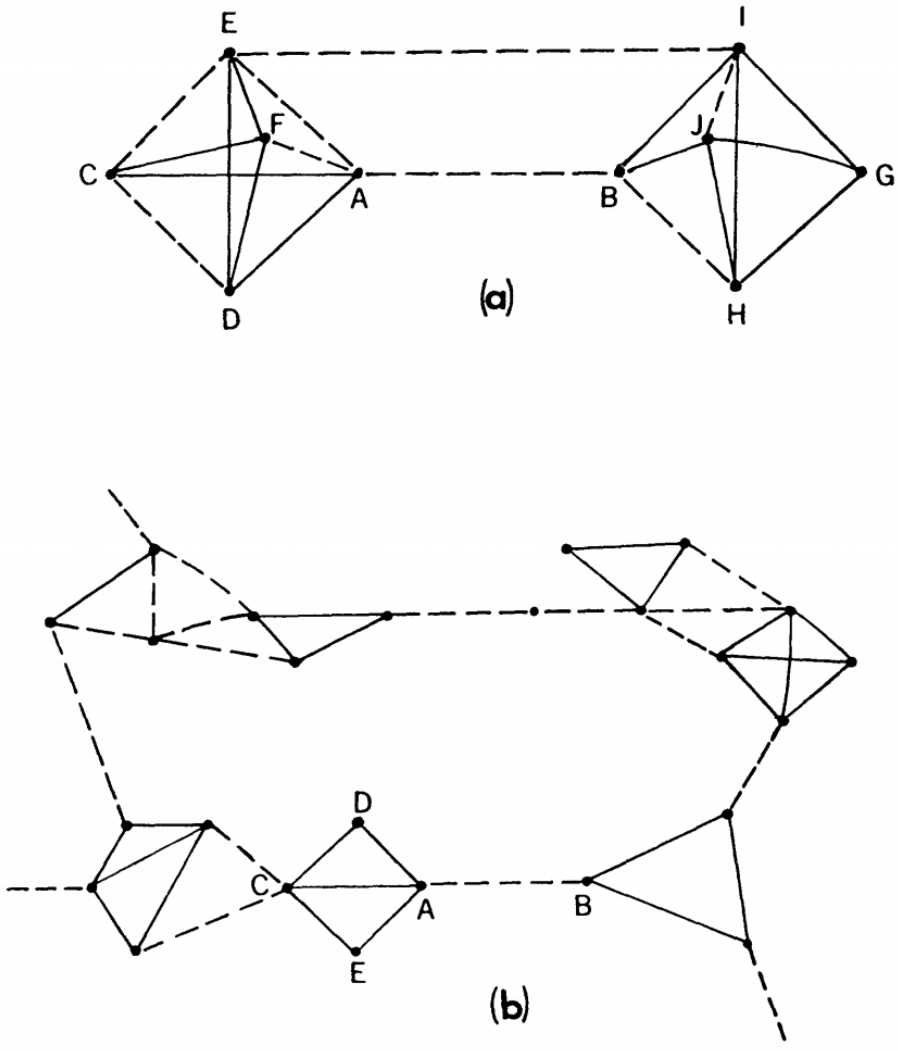

FIG. 2.—Local bridges. *a*, Degree 3; *b*, Degree 13. ——— = strong tie; — — — = weak tie.

Source: Granovetter (1973, AJS):

Digital weak ties

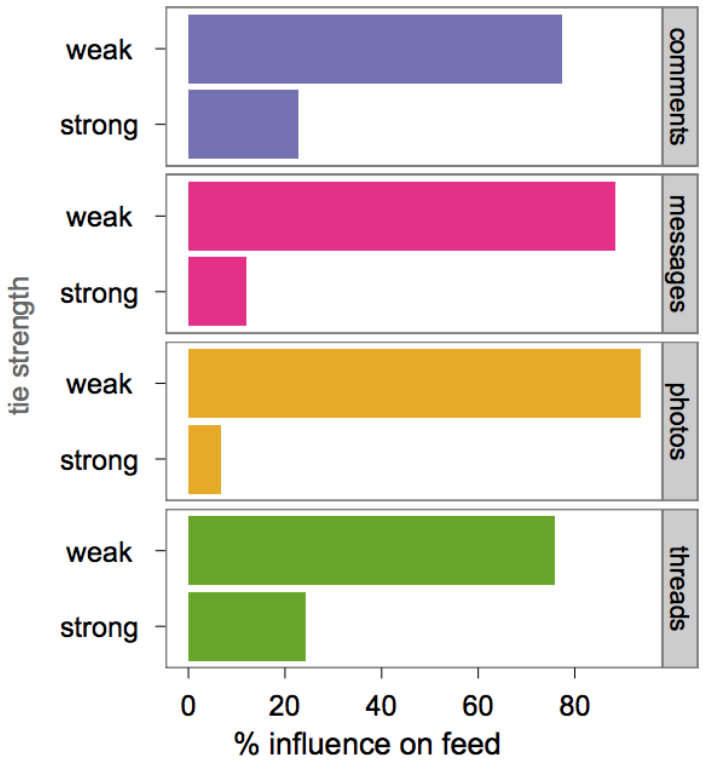

Figure 7: Weak ties are collectively more influential than strong ties. Panels show the percentage of information spread by strong and weak ties for all four measurements of tie strength. Although the probability of influence is significantly higher for those that interact frequently, most contagion occurs along weak ties, which are more abundant.

Bakshy et al (2012):

- Weak ties are responsible for most propagation of novel information on Facebook
- Strong ties provide redundant information
- Suggests contagion processes on Facebook may be more likely to be simple rather than complex

Twitter data

Twitter APIs

Two different methods to collect Twitter data:

1. REST API:
   - Queries for static information about users and tweets
   - Search recent tweets
   - R libraries: tweetscores (also rtweet, academictwitteR)
2. Streaming API:
   - Connect to the "stream" of tweets as they are being published
   - Three streaming APIs:
     - 2.1 Filter stream: tweets filtered by keywords
     - 2.2 Geo stream: tweets filtered by location
     - 2.3 Sample stream: 1% random sample of tweets
   - R library: streamR

Important limitation: *historical* tweets are harder to collect, but now available with some limitations for academic researchers

Anatomy of a tweet

Anatomy of a tweet

Tweets are stored in JSON format:


```
{ "created_at": "Wed Nov 07 04:16:18 +0000 2012",
  "id": 266031293945503744,
  "text": "Four more years. http://t.co/bAJE6Vom",
  "source": "web",
  "user": {
    "id": 813286,
    "name": "Barack Obama",
    "screen_name": "BarackObama",
    "location": "Washington, DC",
    "description": "This account is run by Organizing for Action staff.
        Tweets from the President are signed -bo.",
    "url": "http://t.co/8aJ56Jcemr",
    "protected": false,
    "followers_count": 54873124,
    "friends_count": 654580,
    "listed_count": 202495,
    "created_at": "Mon Mar 05 22:08:25 +0000 2007",
    "time_zone": "Eastern Time (US & Canada)",
    "statuses_count": 10687,
    "lang": "en" },
  "coordinates": null,
  "retweet_count": 756411,
  "favorite_count": 288867,
  "lang": "en"
}
```

Meta data

Datasets:

- Data for good:
  - *Geographic data*: Facebook Population Maps; COVID-19 Mobility Dashboard; Commuting Zones; Climate Conversations; Relative Wealth Index; etc.
  - *Network data*: Social Connectedness Index
  - *Survey data*: Gender Equality; Small Businesses; Climate Change Opinion; COVID Symptoms

- CrowdTangle:
  - Posts on public Facebook Pages and Groups
  - Posts on public Instagram profiles
  - Free access for most academic researchers