Update: Information Diffusion Simulation in the RBUSD Region

Ryan M. Acton

racton@uci.edu

Department of Sociology
University of California, Irvine

April 1, 2008

- The following information goes into producing our simulations:
  - From the decennial U.S. Census (2000):
    - *Population size*—"Actors" are generated and distributed based on the population totals within the 13 Census block groups covering RBUSD. We're working with 63,261 actors, 12 schools, and 1 district office.
    - *Demographic characteristics*—We have information on the marginal distributions of gender, race/ethnicity, age, number of households, . . .

- The following information goes into producing our simulations:
  - From GIS:
    - *Geographic shapefiles*—Gives us geographic boundary coordinates of the region(s) of interest. Allows geographic space to dictate actor placement and inter-actor tie probability
    - *School building coordinates*—We place the schools at their correct geographic locations and "tie" them into the social network of the community. They are the origins of the message.

- Major assumptions inherent to this process (can be tweaked):
  - Actors are distributed uniformly across the landscape (obviously flawed). Can eventually incorporate geographic “holes” in the space to constrain the placement of actors
  - We can only use marginal distributions of demographic characteristics of actors. We know how many of type X people are in each block group, but not which specific people.
  - Assuming all residents within RBUSD’s boundaries are present during the diffusion. Will eventually need to incorporate parts of the population leaving the region to go to work, run errands, etc. at different times of the day

- Major assumptions inherent to this process (can be tweaked):
  - We're basing the information diffusion process on face-to-face communication data—it's our best approximation at present
  - Once you are "infected", you remain infected. You don't "unlearn" information in this scenario.
  - Information diffusion is a serial process (i.e. no two nodes in the network get the information at *exactly* the same time)

- A "seed" node is selected (here, the district office).
- The district office is directly tied to the 12 schools, and those ties are given high diffusion priority
- Each school is presently connected to **one** member of the community. Those ties to community members are also given high diffusion priority.
    - Have we any information about current district policies re: how many (and who) are contacted by each school?
- In terms of priority: District office $\longrightarrow$ schools $\longrightarrow$ community.

- We have a completed diffusion scenario! (Figure 1)
- Computationally it takes $\approx$ 45 minutes to:
  1. Generate the underlying social network among the actors. This simulates how they'd be connected to one another in the course of everyday communication
  2. Run the diffusion routine on the actors in the network

- We can give actors different rates, or propensities, of diffusion
- We can turn links in the network on or off over the course of time
- We've got information on:
    1. Who received the information from whom
    2. The time that a person became "infected" with the information
    3. The "depth" at which one became infected (i.e., how many diffusion steps did it take to reach you?)

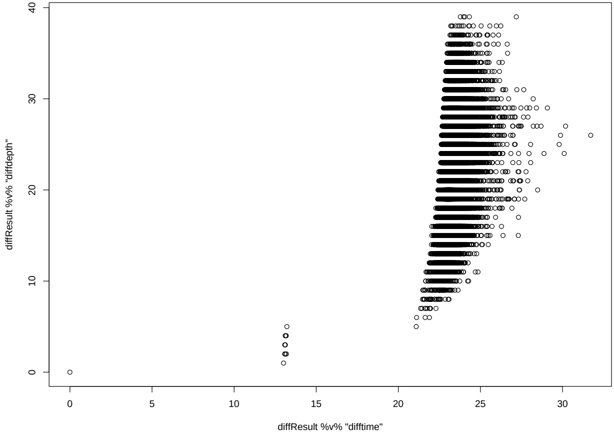

Figure: RBUSD Baseline Diffusion

- The diffusion part is relatively easy. The visualization part is rather complex.
- Ideally would like to animate the diffusion fully—this is the goal for next week
- We want to next employ information about the differential tie probabilities between members of different racial and ethnic groups. Results on this in 1–2 weeks
- Are there any glaring assumptions that need to be better addressed?

Update: Information Diffusion Simulation in the RBUSD Region

Ryan M. Acton

racton@uci.edu

Department of Sociology
University of California, Irvine

April 1, 2008

UCIRVINE | UNIVERSITY of CALIFORNIA