Face to face contact networks

Matthew J. Salganik

Social Network (Soc 204)
Spring 2017
Princeton University

May 1, 2017

Logistics:

- Final class: Online dating

Logistics:

- Final class: Online dating
- Assignment due Wednesday and precept this week

Logistics:

- Final class: Online dating
- Assignment due Wednesday and precept this week
- Calculations on the final: we will provide equations

Logistics:

- Final class: Online dating
- Assignment due Wednesday and precept this week
- Calculations on the final: we will provide equations
- Final exam questions

Partners in crime? Corruption as a criminal network

Romain Feralli

Abstract: This paper explains stylized facts about the organization of bureaucratic corruption across countries using a model of corruption diffusion on a network. Agents are embedded in an organizational network on which they form a criminal subnetwork: an agent finds an illegal rent, and decides whether to spend some of it to recruit accomplices. Accomplices protect the agent but create witnesses among their neighbors, who report corruption. I find that corruption persists and changes form in capable states: it becomes less frequent but selects on cases of grand corruption and involves more accomplices. Second, the relevant unit of analysis is the criminal network, not the individual. Individual accomplices can be highly connected or more isolated, but criminal networks have fewer ties to the organization, and sparser organizations are more corrupt. Depending on the ties, increasing network density may decrease or increase corruption. A lab experiment strongly supports the model's predictions.

Location: Corwin 127, 5/2/2017, 12:00

Breaking your bubble: What are you doing? How is it going? What problems have you encountered?

- P01, W 2:30-3:20, Matt Salganik: **Rahul Mehta**
- P01A, W 2:30-3:20, Romain Ferrali: **Matthew Gancayo**
- P02, W, 3:30-4:20, Romain Ferrali: **Will Rose**
- P03, TH 9:00-9:50, Sarah Reibstein: **Sarah Reibstein**
- P04, TH 10:00-10:50, Sarah Reibstein: **Jessica Nyquist**
- P04A, TH, 10:00-10:50, Sam Clovis: **Grace**
- P04B, TH, 10:00-10:50, Ramina Sotoudeh: **Ramina Sotoudeh**
- P05, TH, 12:30-1:20, Ryan Parsons: **Addie Gilson**
- P05B, TH 12:30-1:20, Ramina Sotoudeh: **Luisa Goytia**
- P06, TH 1:30-2:20, Herrissa Lamothe: **Eric**
- P06A, TH 1:30-2:20 Ramina Sotoudeh: **Gabriela Pitten**
- P07, TH 3:30- 4:20 Herrissa Lamothe: **Grace**
- P07B, TH 3:30-4:20 Janet Xu: **Alice Zheng**
- P10, F 10:00-10:50 Janet Xu: **Matthew and Rachel**

Vote:

1. Mossong et al. (2008). Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases. *PLoS Medicine*.
2. Salathe et al. (2010). A High-Resolution Human Contact Network for Infectious Disease Transmission. *Proceedings of the National Academy of Sciences*.

- Last class: going viral

- Last class: going viral
- This class: really going viral

- Which network is most important for the spread of which disease?

- Which network is most important for the spread of which disease?
- Why should we care about influenza?

How to measure face to face contact networks?

- surveys
- "sociotechnical networks" (e.g., Twitter, Facebook, email, etc)
- mobile phones (e.g., Bluetooth scans)
- wearable sensors

Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases

Joël Mossong[1,2*], Niel Hens[3], Mark Jit[4], Philippe Beutels[5], Kari Auranen[6], Rafael Mikolajczyk[7], Marco Massari[8], Stefania Salmaso[8], Gianpaolo Scalia Tomba[9], Jacco Wallinga[10], Janneke Heijne[10], Malgorzata Sadkowska-Todys[11], Magdalena Rosinska[11], W. John Edmunds[4]

A high-resolution human contact network for infectious disease transmission

Marcel Salathé[a,1,2], Maria Kazandjieva[b], Jung Woo Lee[b], Philip Levis[b], Marcus W. Feldman[a], and James H. Jones[c,d]

Departments of [a]Biology, [b]Computer Sciences, and [c]Anthropology, and [d]Woods Institute for the Environment, Stanford University, Stanford, CA 94305-5020

Edited by Adrian Raftery, University of Washington, Seattle, WA, and approved November 8, 2010 (received for review June 25, 2010)

Compare and contrast

- Settings: 8 European countries vs one American high school

Compare and contrast

- Settings: 8 European countries vs one American high school
- Measurement: diary survey vs sensor

Compare and contrast

- Settings: 8 European countries vs one American high school
- Measurement: diary survey vs sensor
- Contact type: reported contacts/physical contacts and close personal interact

Compare and contrast

- Settings: 8 European countries vs one American high school
- Measurement: diary survey vs sensor
- Contact type: reported contacts/physical contacts and close personal interact
- Network: ego centric vs complete network

Compare and contrast

- Settings: 8 European countries vs one American high school
- Measurement: diary survey vs sensor
- Contact type: reported contacts/physical contacts and close personal interact
- Network: ego centric vs complete network
- Disease outcome of interest: who gets infected first vs vaccination strategies

Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases

Joël Mossong[1,2*], Niel Hens[3], Mark Jit[4], Philippe Beutels[5], Kari Auranen[6], Rafael Mikolajczyk[7], Marco Massari[8], Stefania Salmaso[8], Gianpaolo Scalia Tomba[9], Jacco Wallinga[10], Janneke Heijne[10], Malgorzata Sadkowska-Todys[11], Magdalena Rosinska[11], W. John Edmunds[4]

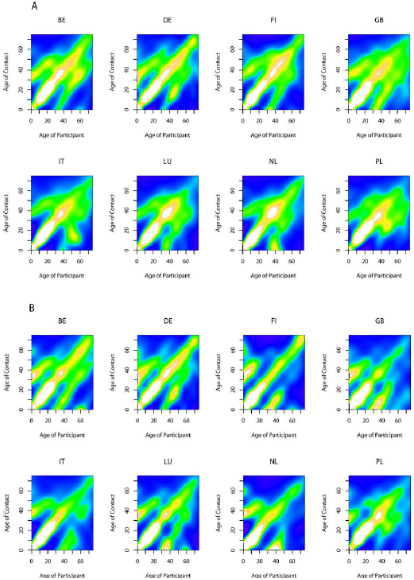


Figure 3. Smoothed Contact Matrices for Each Country Based on (A) All Reported Contacts and (B) Physical Contacts Weighted by Sampling Weights
White indicates high contact rates, green intermediate contact rates, and blue low contact rates, relative to the country-specific contact intensity. Fitting is based on a tensor-product spline to contact matrix data using a negative binomial distribution to account for overdispersion.
doi:10.1371/journal.pmed.0050074.g003

Table 1. Number of Recorded Contacts per Participant per Day by Different Characteristics and Relative Number of Contacts from the Weighted Multiple Censored Negative Binomial Regression Model

Category	Covariate	Number of Participants	Mean (Standard Deviation) of Number of Reported Contacts	Relative Number of Reported Contacts (95% Confidence Interval)[a]
Age of participant, y	0–4	660	10.21 (7.65)	1.00
	5–9	661	14.81 (10.09)	1.42 (1.28–1.55)
	10–14	713	18.22 (12.27)	1.73 (1.57–1.90)
	15–19	685	17.58 (12.03)	1.68 (1.52–1.84)
	20–29	879	13.57 (10.60)	1.45 (1.33–1.57)
	30–39	815	14.14 (10.15)	1.45 (1.34–1.57)
	40–49	908	13.83 (10.86)	1.38 (1.27–1.50)
	50–59	906	12.30 (10.23)	1.31 (1.20–1.42)
	60–69	728	9.21 (7.96)	1.06 (0.96–1.16)
	70+	270	6.89 (5.83)	0.81 (0.73–0.88)
	Missing value	65	9.63 (9.05)	0.91 (0.66–1.17)
Sex of participant	Female	3,808	13.39 (10.57)	1.00
	Male	3429	13.51 (10.67)	0.99 (0.96–1.02)
	Missing value	53	10.92 (8.60)	1.57 (1.09–2.05)
Household size	1	749	8.87 (8.27)	1.00
	2	1,645	10.65 (9.14)	1.17 (1.11–1.24)
	3	1,683	12.87 (10.26)	1.20 (1.13–1.27)
	4	2,041	15.84 (11.17)	1.36 (1.28–1.44)
	5	814	16.47 (11.21)	1.46 (1.35–1.56)
	6+	358	17.69 (10.98)	1.56 (1.43–1.70)
Day of the week	Sunday	862	10.10 (8.76)	1.00
	Monday	1,032	13.32 (10.31)	1.33 (1.24–1.41)
	Tuesday	1,116	14.17 (10.83)	1.39 (1.31–1.48)
	Wednesday	1,017	14.58 (11.14)	1.38 (1.29–1.47)
	Thursday	1,069	14.70 (11.23)	1.41 (1.32–1.50)
	Friday	1,122	14.72 (11.25)	1.43 (1.34–1.52)
	Saturday	936	11.63 (9.11)	1.20 (1.12–1.28)
	Missing value	136	12.48 (10.66)	1.24 (1.08–1.40)
Country[b]	BE	750	11.84 (9.85)	1.00
	DE	1,341	7.95 (6.26)	0.70 (0.65–0.74)
	FI	1,006	11.06 (7.89)	0.94 (0.88–1.00)
	GB	1,012	11.74 (7.67)	0.99 (0.92–1.05)
	IT	849	19.77 (12.27)	1.66 (1.55–1.78)
	LU	1,051	17.46 (12.81)	1.42 (1.33–1.51)
	NL	269	13.85 (10.54)	1.34 (1.20–1.47)
	PL	1,012	16.31 (11.45)	1.37 (1.28–1.47)

[a]Dispersion parameter alpha = 0.36 (95% CI 0.34–0.37); alpha = 0 would correspond to no overdispersion, i.e., a censored Poisson distribution.
[b]Direct comparisons between countries are difficult because of different approaches to recording frequent professional contacts. In BE, DE, FI, and NL, participants were instructed not to record professional contacts in the diary if they had more than 20 (BE) or 10 (DE, FI, NL) of them per day.
doi:10.1371/journal.pmed.0050074.t001

Note different rates of contact in different countries

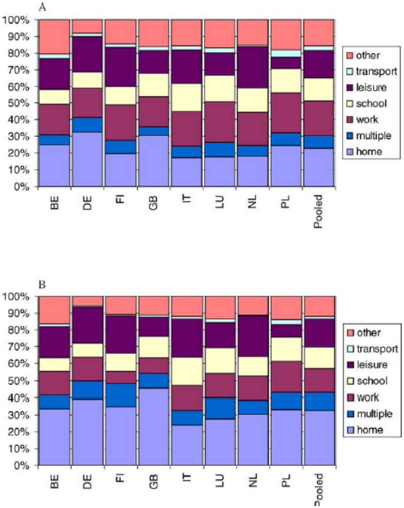


Figure 2. The Distribution by Location and by Country of (A) All Reported Contacts and (B) Physical Contacts Only

Sampling weights were used for each country. "Other" refers to contacts made at locations other than home, work, school, travel, or leisure. "Multiple" refers to the fact that the person was contacted during the day in multiple locations, not just a single location.

doi:10.1371/journal.pmed.0050074.g002

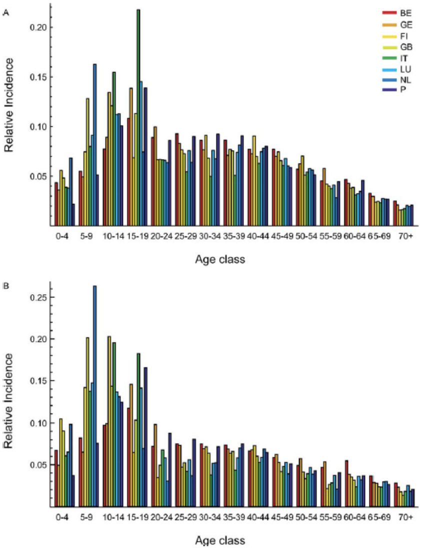


Figure 4. Relative Incidence of a New Emerging Infection in a Completely Susceptible Population, When the Infection Is Spread between and within Age Groups by the Contacts as Observed in Figure 3

For each country, we monitored incidence five generations of infection after the introduction of a single infected individual in the 65–70 age group; the incidence is normalized such that height of all bars sums to one for each country. (A) Results for all reported contacts; (B) for physical contacts only.

doi:10.1371/journal.pmed.0050074.g004

A high-resolution human contact network for infectious disease transmission

Marcel Salathé[a,1,2], Maria Kazandjieva[b], Jung Woo Lee[b], Philip Levis[b], Marcus W. Feldman[a], and James H. Jones[c,d]

Departments of [a]Biology, [b]Computer Sciences, and [c]Anthropology, and [d]Woods Institute for the Environment, Stanford University, Stanford, CA 94305-5020

Edited by Adrian Raftery, University of Washington, Seattle, WA, and approved November 8, 2010 (received for review June 25, 2010)

Converting interactions to networks

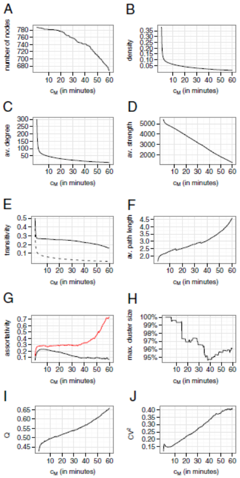
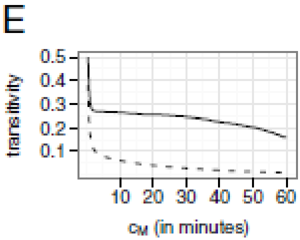
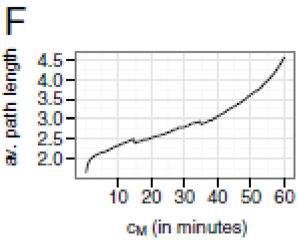


Fig. 2. Various statistics on the contact graph with minimum contact duration, c_m (i.e., the left-most point in each panel represents the full contact graph, the right-most point represents the contact graph that contains only contacts that are at least 60 min long). With increasing c_m, nodes drop out of the network if they have no contact that satisfies the minimum duration condition. (*A*) Hence, the reduction in the number, V, of nodes. (*B*) Density of the graph ($2E/[V*(V-1)]$), where E is the number of edges. (*C*) Average (av.) degree. (*D*) av. strength, where the strength of a node is the total number of CPRs of the node. (*E*) Transitivity (i.e., cluster coefficient) as defined by Barrat et al. (25) and expected value (mean degree/V) in a random network (dashed line). (*F*) Average path length. (*G*) Assortativity (23) with respect to degree (black line) and role (red line). (*H*) Size of the largest component as a fraction of total network size. max., maximum. (*I*) Modularity, Q, as defined by Reichardt and Bornholdt (39). (*J*) CV^2 of degree.


E
0.5
0.4
0.3
0.2
0.1
transitivity
10 20 30 40 50 60
cM (in minutes)

F
4.5
4.0
3.5
3.0
2.5
2.0
av. path length
10 20 30 40 50 60
cM (in minutes)

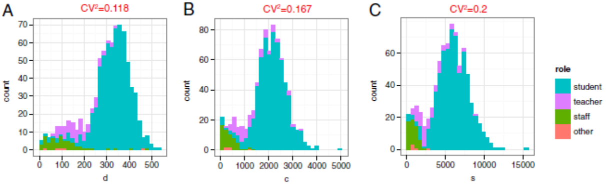


Fig. 3. Distribution and CV^2 of degree, d (*A*); number of interactions, c (*B*); and strength, s (*C*), based on the full contact network and colored by the role of individuals.

Not power law!

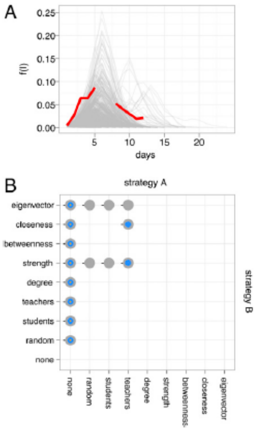


Fig. 4. (*A*) Absentee data (red) and data generated by the SEIR model (gray; 1,000 runs with $R_0 > 1$ shown). Gray lines show frequency of infectious individuals, f(I); red lines show the combined frequency of students who reported, or were diagnosed with, a fever and teachers who were absent (gap in the line attributable to weekend). (*B*) Differences in effect of vaccination strategies. Colors represent vaccination coverage of 5% (orange), 10% (blue), and 20% (gray). A point at the intersection of strategy A and strategy B indicated that between those strategies, there was a significant difference ($P < 0.05$, two-sided Wilcoxon test) in the outbreak size at all transmission probability values at the given vaccination coverage. A black horizontal or vertical line points in the direction of the strategy that resulted in smaller outbreak sizes. Because of the symmetry of the grid, data points below the left bottom and top right diagonal line are not shown.

Summary:

- influenza (and other diseases) spread through face to face contact

Summary:

- influenza (and other diseases) spread through face to face contact
- measuring face to face contact networks is difficult and we don't know the best way to do it

Summary:

- influenza (and other diseases) spread through face to face contact
- measuring face to face contact networks is difficult and we don't know the best way to do it
- could be useful for epidemiology and sociology

http://bit.ly/socnet204

http://bit.ly/socnet204

Online dating

- Effect of digital system on society
- Using digital system to study society