"The secret of my influence has always been that it remained secret."

– Salvador Dalí

Gossip: Identifying central individuals in a social network

Abhijit Banerjee* Arun G. Chandrasekhar‡
Esther Duflo* Matthew Jackson‡

*MIT, ‡Stanford

MOTIVATION

- Understanding who is influential in a community is important
  - for members of a network
  - for people trying to sell new things
  - for policymakers who seek to leverage the community to maximize the reach of training, support public health campaign etc.
- Problem is: who is influential?
  - Network theory suggests ideas: people with high centrality.
  - Research suggests that shortcuts may not work (people with many friends, people who live in central location).
- How about asking a few people in the network? perhaps surprisingly, this is not a suggestion in the literature (in economics or marketing). We do know that community members are good at identifying the poor (Alatas et al), or the productive (Hussam et al). Yet far from automatic that members of a social network should know who is central
  - hard to know how central others are outside of immediate circle
  - research suggests network members have poor image of network, beyond immediate friends

CONTRIBUTIONS

1. "Pure" information experiment, spreading information on a lottery for cell phone and cash prizes.
   - Randomized experiment in 213 villages: randomize villages to seeding gossip nominees (who is good to spread information about a fair or a promotion), elders or random
   - Seeding information with "gossip nominees" triples info spread
2. Much more relevant policy context: spread of immunization in Haryana
   - 516 villages were seeded information on immunization
   - random, Trusted, "Gossip" or Trusted Gossip.
   - Gossip increase number of kids immunized for all different shots by 18%

CONTRIBUTIONS

1. "Pure" information experiment, spreading information on a lottery for cell phone and cash prizes.
    - Randomized experiment in 213 villages: randomize villages to seeding gossip nominees (who is good to spread information about a fair or a promotion), elders or random
    - Seeding information with "gossip nominees" triples info spread
2. Much more relevant policy context: spread of immunization in Haryana
    - 516 villages were seeded information on immunization
    - random, Trusted, "Gossip" or Trusted Gossip.
    - Gossip increase number of kids immunized for all different shots by 18%
3. Understanding Channel: Model characterizing **why** and **how** boundedly rational members of a network could (easily) know who is diffusion central, by tracking "gossip".
    - Even if agents know nothing about network structure

CONTRIBUTIONS

1. "Pure" information experiment, spreading information on a lottery for cell phone and cash prizes.
    - Randomized experiment in 213 villages: randomize villages to seeding gossip nominees (who is good to spread information about a fair or a promotion), elders or random
    - Seeding information with "gossip nominees" triples info spread
2. Much more relevant policy context: spread of immunization in Haryana
    - 516 villages were seeded information on immunization
    - random, Trusted, "Gossip" or Trusted Gossip.
    - Gossip increase number of kids immunized for all different shots by 18%
3. Understanding Channel: Model characterizing **why** and **how** boundedly rational members of a network could (easily) know who is diffusion central, by tracking "gossip".
    - Even if agents know nothing about network structure
4. Exploring Channels empirically:
    - Using Wave II network data from BCDJ (2013), 35 networks: is diffusion centrality a better predictor than other X's?
    - Using experiment 1 combined with full network data: does diffusion centrality fully explain the results?

Contributions

1. "Pure" information experiment, spreading information on a lottery for cell phone and cash prizes.
   - Randomized experiment in 213 villages: randomize villages to seeding gossip nominees (who is good to spread information about a fair or a promotion), elders or random
   - Seeding information with "gossip nominees" triples info spread
2. Much more relevant policy context: spread of immunization in Haryana
   - 516 villages were seeded information on immunization
   - random, Trusted, "Gossip" or Trusted Gossip.
   - Gossip increase number of kids immunized for all different shots by 18%
3. Understanding Channel: Model characterizing **why** and **how** boundedly rational members of a network could (easily) know who is diffusion central, by tracking "gossip".
   - Even if agents know nothing about network structure
4. Exploring Channels empirically:
   - Using Wave II network data from BCDJ (2013), 35 networks: is diffusion centrality a better predictor than other X's?
   - Using experiment 1 combined with full network data: does diffusion centrality fully explain the results?

experiment 1: Diffusing simple information

DESIGN

- setting: 213 new villages in Karnataka, India
- partner: cellphone sales company
- in each village one week before the experiment
  - (attempt to) contact 10-15 random households
  - households administered gossip question
  - households asked to name village elders
- information to be diffused:
  - a promotion wherein if households give us a missed call, they have the opportunity to win a new phone or cash prizes
  - lottery:
    - roll two 6's: win a phone
    - other rolls: correspond to cash prizes

DESIGN

In each village, tell $k \in \{3, 5\}$ seeds about the promotion.

213 villages across 3 treatments (71 villages / treatment).

T1. **Village Elders:**
- k randomly chosen from list obtained 1 week prior

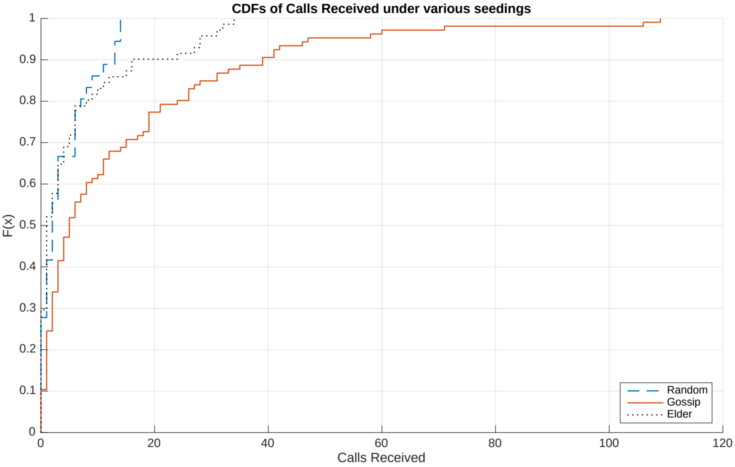
T2. **Random:**
- households chosen uniformly at random via right-hand-rule
- sometimes hit gossips by chance

T3. **Gossips:**
- k randomly chosen from list obtained 1 week prior

Outcome: Number of missed calls we receive
specification: Intent to treat; OLS of hitting at least one gossip; IV of hitting at least one gossip.

Num. calls regressed on treatment

VARIABLES	(1) RF Calls Received	(2) OLS Calls Received	(3) IV Calls Received	(4) RF Calls Received Seeds	(5) OLS Calls Received Seeds	(6) IV Calls Received Seeds
Gossip Treatment	4.261			1.166		
	(2.672)			(0.666)		
Elder Treatment	-2.922	-0.103	0.667	-0.605	-0.0714	0.377
	(1.723)	(3.082)	(3.270)	(0.444)	(0.477)	(0.823)
At least 1 Gossip		6.420	7.850		1.357	2.148
		(2.210)	(4.852)		(0.544)	(1.218)
Constant	8.443	6.293	4.854	2.092	2.008	1.111
	(1.369)	(2.333)	(3.101)	(0.323)	(0.545)	(0.766)
Observations	212	212	212	212	212	212


CDFs of Calls Received under various seedings
F(x)
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
20
40
60
80
100
120
Calls Received
Random
Gossip
Elder

experiment 2: Diffusing A Message about immunization

Background

- Large scale project undertaken in collaboration with the government of Haryana, India
- Objective is to increase demand for immunization [in a context with low immunization rate]
- We developed and deployed in a large sample of villages a e-health application on Android. Serves as set up for several experiments:
  - Incentives
  - Reminders
  - "Seed" intervention

Design of the Seed intervention

- (well) before the tablets and incentive treatment started, we visited 516 villages in the experiment and ask random households to nominate up to 4 people.
- Villages were randomly selected to be:
  1. Gossip
  2. Trusted
  3. Trusted Gossip
- Then we selected randomly 5 of the people who had been nominated, and enrolled them as seed

"Who are the people in this village, who when they share information, many people in the village get to know about it. For example, if they share information about a music festival, street play, fair in this village, or movie shooting many people would learn about it. This is because they have a wide network of friends/contacts in the village and they can use that to actively spread information to many villagers. Could you name four such individuals, male or female, that live in the village (within OR outside your neighborhood in the village) who when they say something many people get to know?"

Trusted

"Who are the people in this village that you and many villagers trust, both within and outside this neighborhood, trust? When I say trust I mean that when they give advice on something, many people believe that it is correct and tend to follow it. This could be advice on anything like choosing the right fertilizer for your crops, or keeping your child healthy. Could you name four such individuals, male or female, who live in the village (within OR outside your neighborhood in the village) and are trusted?"

"Who are the people in this village, both within and outside this neighborhood, who when they share information, many people in the village get to know about it. For example, if they share information about a music festival, street play, fair in this village, or movie shooting many people would learn about it. This is because they have a wide network of friends/contacts in the village and they can use that to actively spread information to many villagers. Among these people, who are the people that you and many villagers trust? When I say trust I mean that when they give advice on something, many people believe that it is correct and tend to follow it. This could be advice on anything like choosing the right fertilizer for your crops, or keeping your child healthy. Could you name four such individuals, male or female, that live in the village (within OR outside your neighborhood in the village) who when they say something many people get to know and are trusted by you and other villagers?"

Intervention

- We visited them once before anything else to get their consent to get messaged once a month. We told them about the importance of immunization and suggest they spread it.
- From then on, we messaged them once a month with text messages that say the following:
  - In Incentive villages: Vaccination protects your child from 10 types of diseases and ensures complete physical and mental development of the child . Families with children below 12 months of age will receive a free mobile recharge worth TK as a gift for vaccinating their child. Please share this information with your friends and family members and encourage them to immunize their child at the nearest immunization session camp.
  - In No incentive villages: Vaccination protects your child from 10 types of diseases and ensures complete physical and mental development of the child . Please share this information with your friends and family members and encourage them to immunize their child at the nearest session camp.

RESULTS

	Log(Number of Children received Penta1)	Log(Number of Children received Penta2)	Log(Number of Children received Penta3)	Log(Number of Children received Measles)
	(1)	(2)	(3)	(4)
Gossip	0.146	0.192	0.190	0.181
	(0.100)	(0.097)	(0.094)	(0.086)
Trusted	0.141	0.159	0.149	0.119
	(0.092)	(0.088)	(0.088)	(0.083)
Trusted Gossip	0.129	0.146	0.178	0.124
	(0.093)	(0.089)	(0.086)	(0.078)
Slope	0.135	0.143	0.159	0.148
	(0.085)	(0.082)	(0.081)	(0.074)
Flat	-0.013	0.025	0.085	0.044
	(0.098)	(0.096)	(0.090)	(0.084)
Control Mean	9.02	7.43	6.35	4.37
Observations (village x month)	3,543	3,468	3,406	3,175

Slope incentive only

	Log(Number of Children received Penta1)	Log(Number of Children received Penta2)	Log(Number of Children received Penta3)	Log(Number of Children received Measles)
	(1)	(2)	(3)	(4)
Gossip	0.078	0.115	0.164	0.195
	(0.184)	(0.178)	(0.176)	(0.162)
Trusted	0.254	0.245	0.230	0.194
	(0.167)	(0.164)	(0.164)	(0.161)
Trusted Gossip	0.088	0.110	0.172	0.112
	(0.166)	(0.160)	(0.160)	(0.144)
Control Mean	9.02	7.43	6.35	4.37
Observations (village x month)	1,117	1,096	1,071	1,007

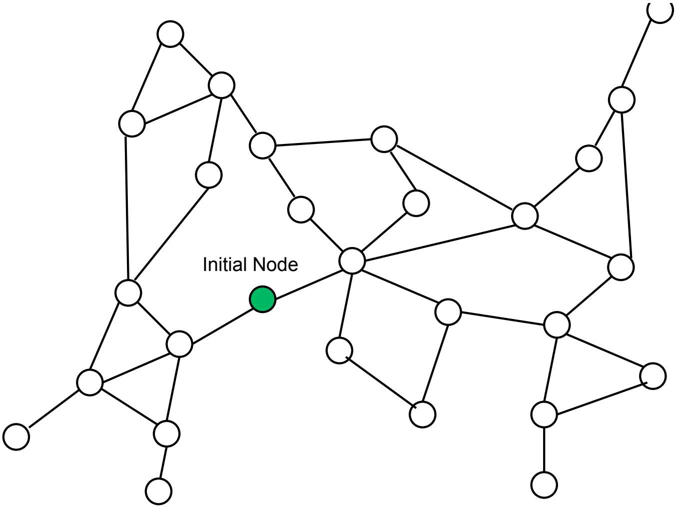
Flat incentive or no incentive

	Log(Number of Children received Penta1)	Log(Number of Children received Penta2)	Log(Number of Children received Penta3)	Log(Number of Children received Measles)
	(1)	(2)	(3)	(4)
Gossip	0.150	0.208	0.183	0.145
	(0.112)	(0.110)	(0.106)	(0.093)
Trusted	0.054	0.093	0.085	0.052
	(0.111)	(0.107)	(0.108)	(0.097)
Trusted Gossip	0.124	0.143	0.160	0.110
	(0.108)	(0.105)	(0.099)	(0.089)
Control Mean	9.02	7.43	6.35	4.37
Observations (village x month)	2,426	2,372	2,335	2,168

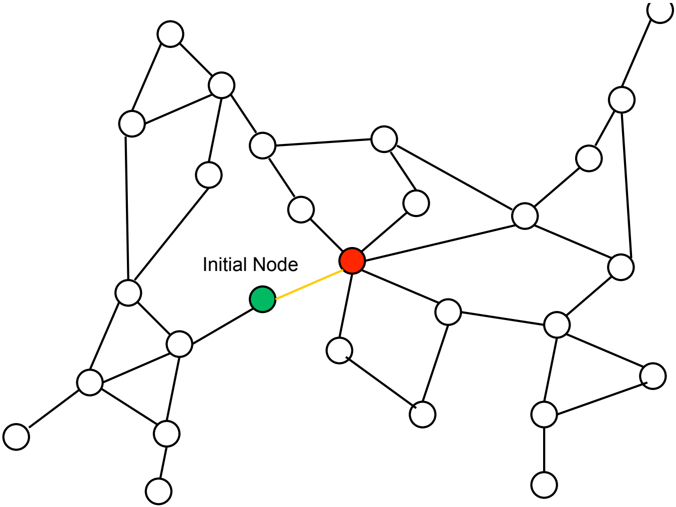
Why would people be able to identify those who are good at communicating information?

Information diffusion: 4 periods, probability of passing=0.5

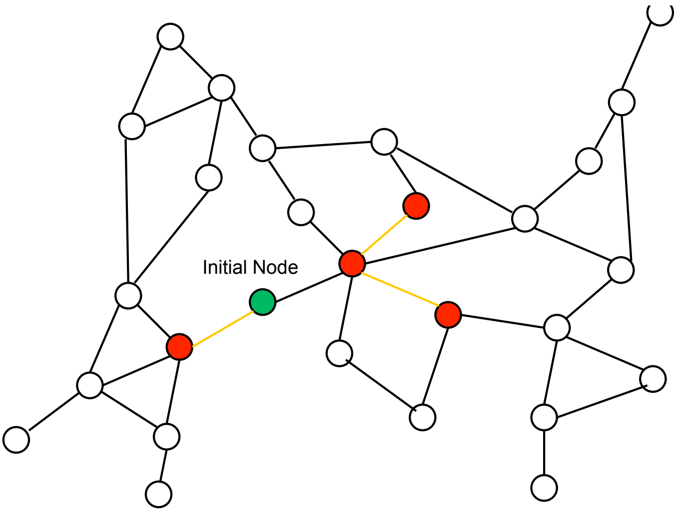
DIFFUSION CENTRALITY: $DC_i(0.5, 4)$



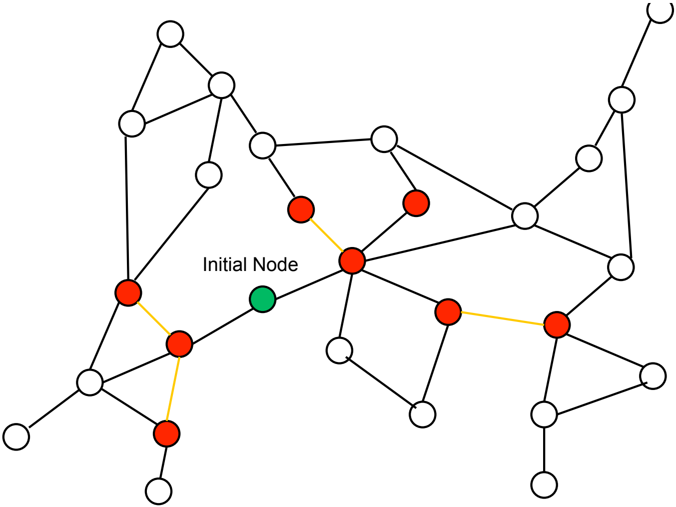
diffusion centrality: $DC_i(0.5, 4)$



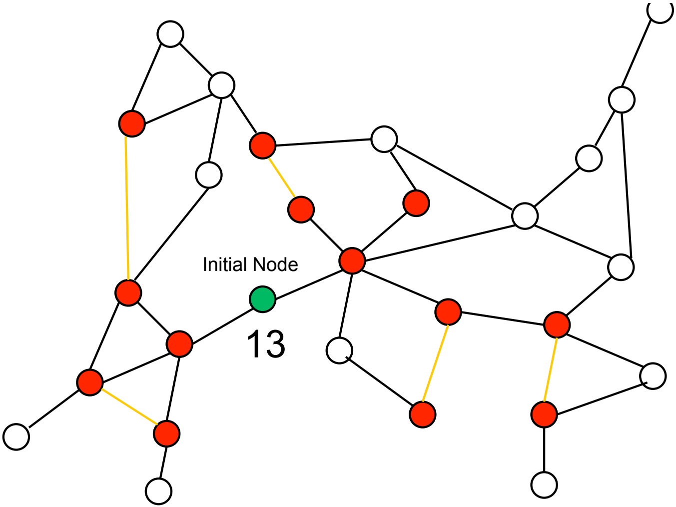
DIFFUSION CENTRALITY: $DC_i(0.5, 4)$



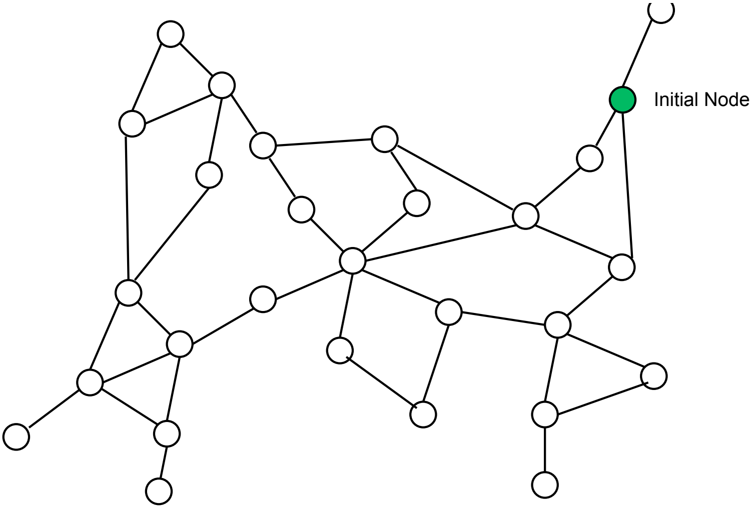
DIFFUSION CENTRALITY: $DC_i(0.5, 4)$



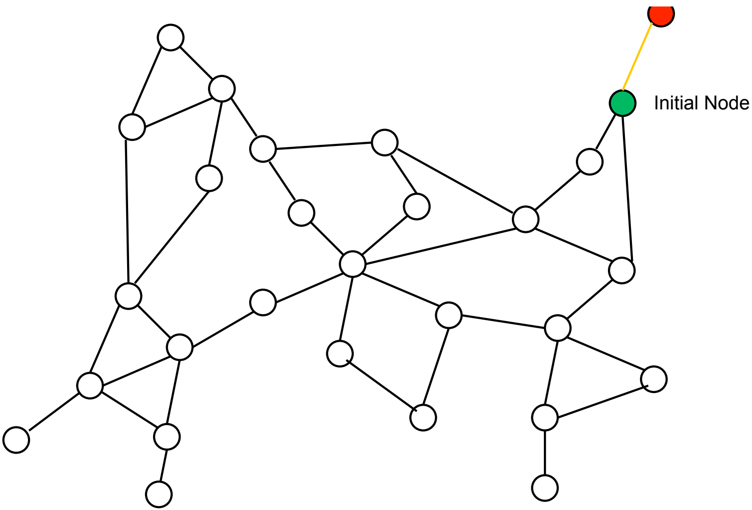
diffusion centrality: $DC_i(0.5, 4)$



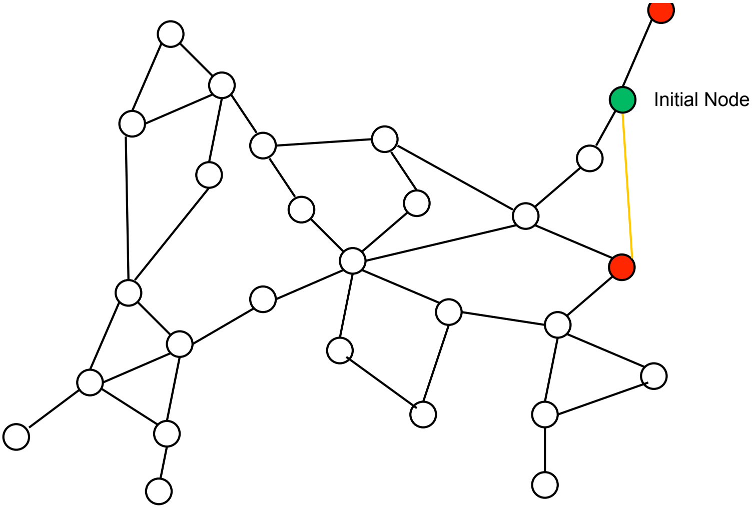
diffusion centrality: $DC_i(0.5, 4)$



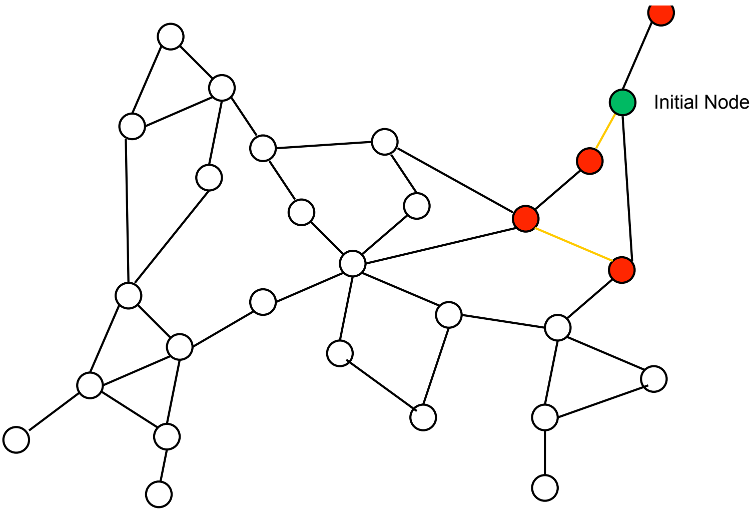
DIFFUSION CENTRALITY: $DC_i(0.5, 4)$



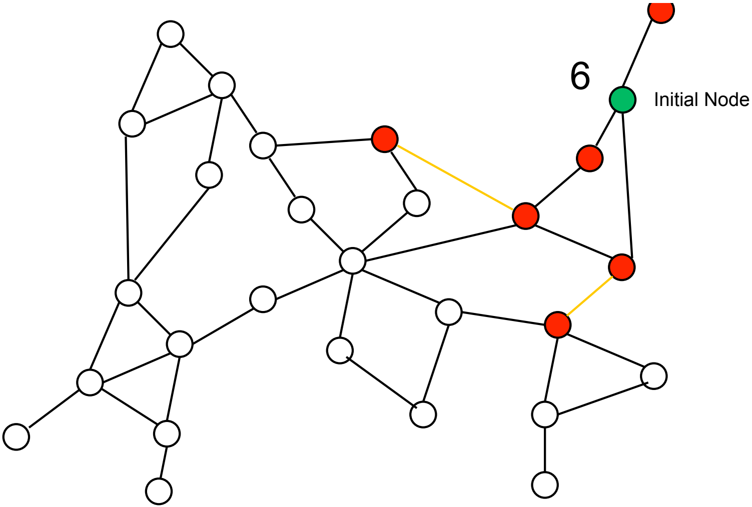
DIFFUSION CENTRALITY: $DC_i(0.5, 4)$



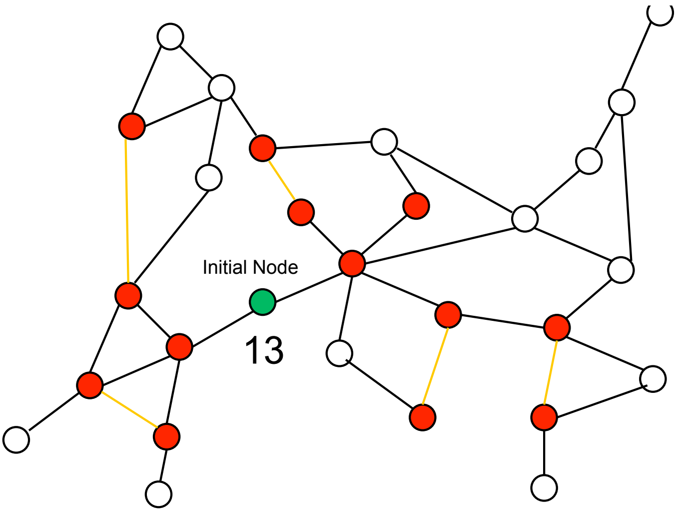
DIFFUSION CENTRALITY: $DC_i(0.5, 4)$



DIFFUSION CENTRALITY: $DC_i(0.5, 4)$



DIFFUSION CENTRALITY: $DC_i(0.5, 4)$



$$DC(\mathbf{g}; q, T) := \left(\sum_{t=1}^{T} (q\mathbf{g})^t \right) \cdot \mathbf{1}.$$

- DC_i is the total expected number of times information starting at i hits all others

What is the relationship between diffusion centrality and others?

Theorem:

1. Diffusion centrality is proportional to degree when $T = 1$:

$$DC(\mathbf{g}; q, 1) = qd(\mathbf{g}).$$

Theorem:

1. Diffusion centrality is proportional to degree when $T = 1$:

$$DC\left(\mathbf{g}; q, 1\right) = qd\left(\mathbf{g}\right).$$

2. If $\mathbf{g}$ is irreducible and aperiodic and $q \geq 1/\lambda_1$, then as $T \to \infty$ diffusion centrality approximates eigenvector centrality:

$$\lim_{T\to\infty} \frac{1}{\sum_{t=1}^{T}\left(q\lambda_1\right)^t} DC\left(\mathbf{g}; q, T\right) = e\left(\mathbf{g}\right).$$

Theorem:

1. Diffusion centrality is proportional to degree when $T = 1$:

$$DC\left(\mathbf{g}; q, 1\right) = qd\left(\mathbf{g}\right).$$

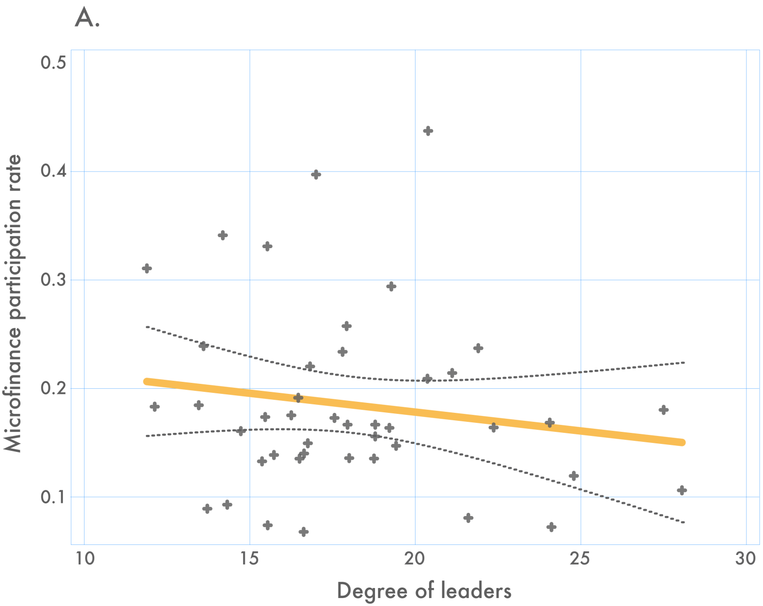
2. If $\mathbf{g}$ is irreducible and aperiodic and $q \geq 1/\lambda_1$, then as $T \to \infty$ diffusion centrality approximates eigenvector centrality:

$$\lim_{T\to\infty} \frac{1}{\sum_{t=1}^{T} \left(q\lambda_1\right)^t} DC\left(\mathbf{g}; q, T\right) = e\left(\mathbf{g}\right).$$

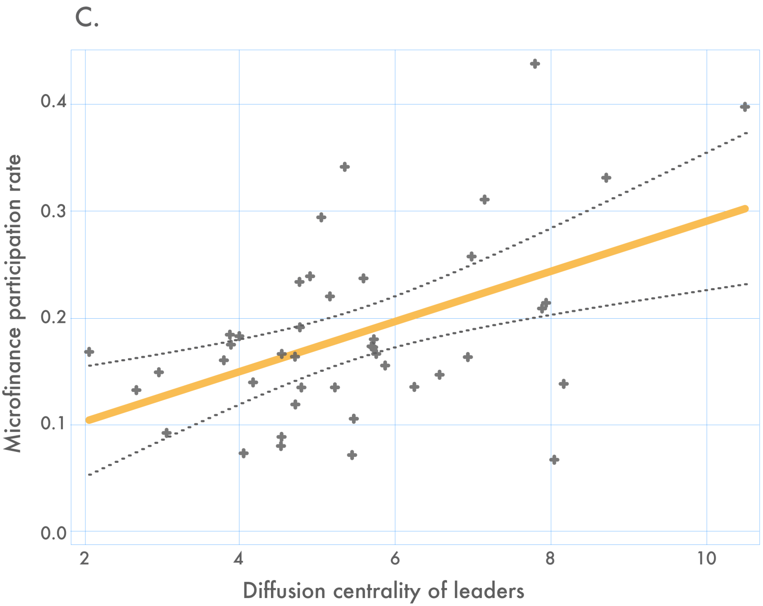
3. For $T = \infty$ and $q < 1/\lambda_1$, diffusion centrality is Katz-Bonacich centrality:

$$DC\left(\mathbf{g}; q, \infty\right) = KB\left(\mathbf{g}, q\right); \quad q < 1/\lambda_1.$$

DEGREE DOESN'T PREDICT TAKE-UP



Diffusion centrality does



Can network member identify those with high DC?

- BCDJ ('13) and Beaman et al. ('14) show value of hitting injection point with high EV centrality
- But it is very expensive to collect network data
  - not a scalable policy solution
- What about asking members of the network?
- Ex ante, should not expect people to know who may be central
  - eigenvector centrality depends on macro-structure of the network
  - people are bad at knowing network structure at arms length (Carley and Krackhardt, 1996; Krackhardt and Kilduff, 1999; Breza, Chandrasekhar, and Tahbaz-Salehi, 2016)
- On the other hand, if they knew, they could also know other things about these people that makes them even better to pass some type of information: they could do even better

compare the process a from how listeners rank others (gossip centrality) to that from the sender's perspective (diffusion centrality)

- “Matt changed jobs”, “Esther bought a goat,” spreads randomly
- Probability q that news is passed from one node to another
- Keep track of how many times hear news about Matt, Esther...

Let

$$\mathbf{M}(\mathbf{g}; q, T) := \left(\sum_{t=1}^{T} (q\mathbf{g})^t \right).$$

$M(\mathbf{g}; q, T)_{ij}$ is expected number of times j hears a piece of information originating from i.

Let

$$\mathbf{M}(\mathbf{g}; q, T) := \left(\sum_{t=1}^{T} (q\mathbf{g})^t \right).$$

$M(\mathbf{g}; q, T)_{ij}$ is expected number of times j hears a piece of information originating from i.

Define *network gossip heard* by node j to be

$$NG(\mathbf{g}; q, T)_j = M(\mathbf{g}; q, T)_{\cdot j}.$$

Let

$$\mathbf{M}(\mathbf{g};q,T) := \left(\sum_{t=1}^{T}(q\mathbf{g})^t\right).$$

$M(\mathbf{g};q,T)_{ij}$ is expected number of times j hears a piece of information originating from i.

Define *network gossip heard* by node j to be

$$NG(\mathbf{g};q,T)_j = M(\mathbf{g};q,T)_{\cdot j}.$$

Conceptual difference:

- Diffusion centrality tracks how well info spreads from a given node
- Network gossip tracks how relatively often j hears about info originating from other nodes

Every individual's rankings of others under network gossip will be **according to the ranking of diffusion centrality** for large enough T and q.

Theorem: If $\mathbf{g}$ is irreducible and aperiodic, and if $q \geq 1/\lambda_1$, then as $T \to \infty$

- every individual j's ranking of others under $NG(\mathbf{g}; q, T)_j$ will be according to the ranking of diffusion centrality, $DC(\mathbf{g}; q, T)$,
- and hence according to eigenvector centrality, $e(\mathbf{g})$.

Intuition:

- much more likely to hear about a central node's gossip relative to a nearby, non-central friend with enough communication periods.

finite T?

- For large amounts of communication, everyone can rank everyone else's diffusion centrality/eigenvector centrality correctly
- Are people still correct, at least on average, for small T?

Theorem: For any $(\mathbf{g}; p, T)$,

$$\sum_j \mathrm{cov}(DC(\mathbf{g}; p, T), NG(\mathbf{g}; p, T)_j) = \mathrm{var}(DC).$$

So, generally, there is a positive correlation between diffusion centrality of i and how much i is heard about by different j's, regardless of T

gossip leads to accurate network knowledge

- With many iterations, all people's rankings of others based on gossip will match diffusion centrality/eigenvector centrality
- For small numbers of iterations, the average network gossip correlates positively with diffusion centrality

EVIDENCE

Ask every adult who is best staring point for diffusion

- Work with households as units
- 35 villages, avg 196 households per village
- Detailed network data on 12+ dimensions of interactions for 90% of households in every village (BCDJ, Wave II)

Eliciting centrality

1. *"If we want to spread information about a new loan product to everyone in your village to whom do you suggest we speak?"*

Eliciting centrality

1. *"If we want to spread information about a new loan product to everyone in your village to whom do you suggest we speak?"*

2. *"If we want to spread information to everyone in the village about tickets to a music event, drama, or fair that we would like to organize in your village, to whom should we speak?"*

HOW WELL DO PEOPLE DO

- People focus on a few people.
- Do they name highly central people?

- People focus on a few people.
- Do they name highly central people?

FOCUS ON A FEW

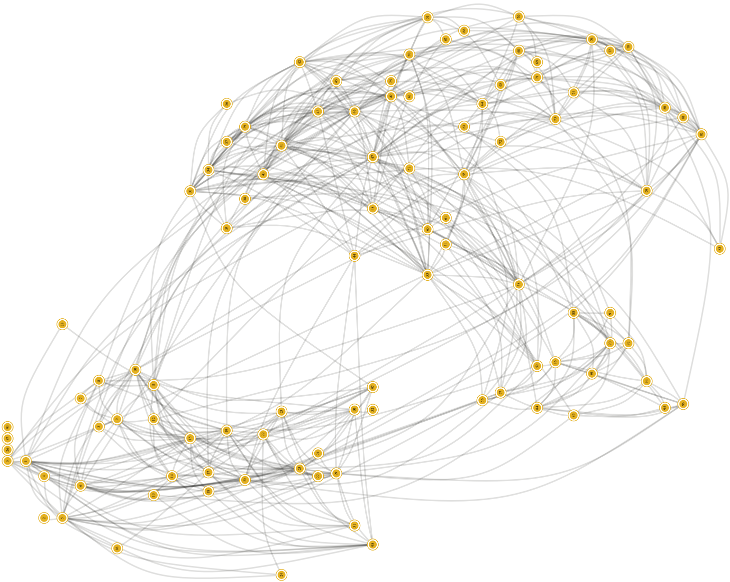
	mean	sd
households per village	196	61.70
household degree	17.72	(9.81)
clustering in a household's neighborhood	0.29	(0.16)
avg distance between nodes in a village	2.37	(0.33)
fraction in the giant component	0.98	(0.01)
is a "leader"	0.13	(0.34)
nominated someone for event	0.38	(0.16)
nominated someone for loan	0.48	(0.16)
was nominated for event	0.04	(0.02)
was nominated for loan	0.05	(0.03)
number of nominations received for loan	0.45	(3.91)
number of nominations received for event	0.34	(3.28)

- People focus on a few people.
- Do they name highly central people?

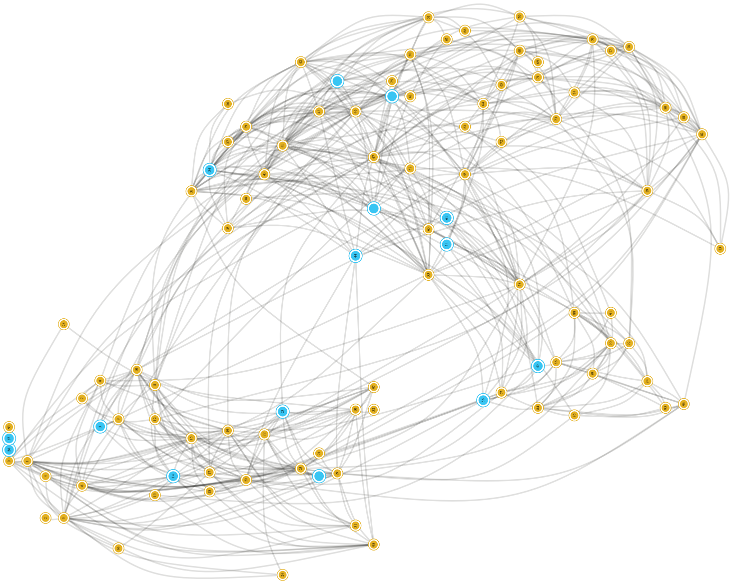
HOW WELL DO PEOPLE DO

- People focus on a few people.
- Do they name highly central people?

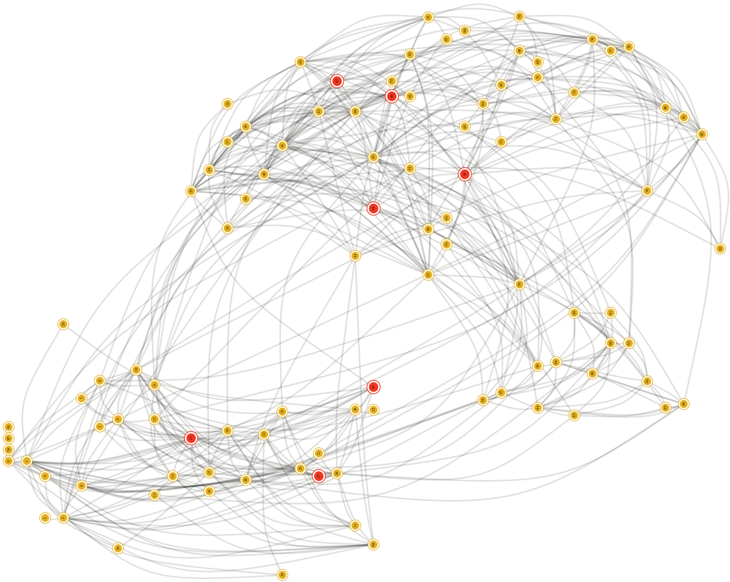
VILLAGE



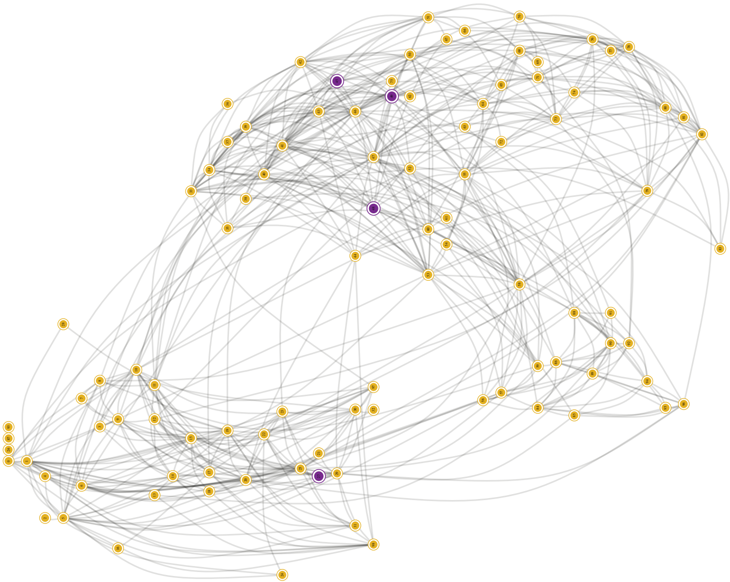
LEADERS (MFI DEFINITION)



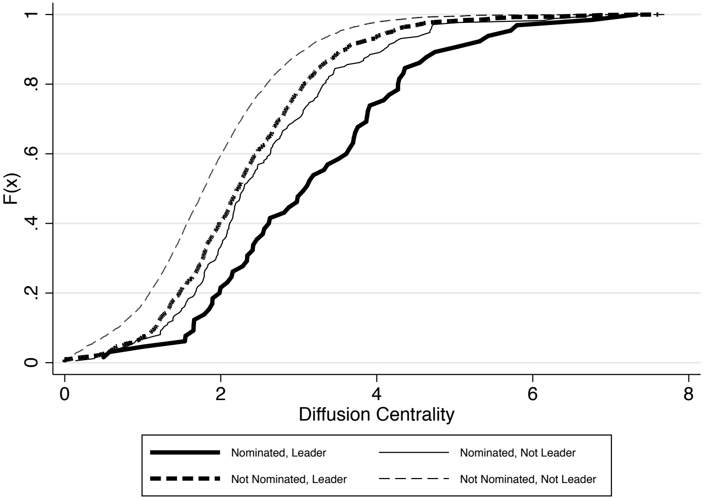
GOSSIP NOMINEES



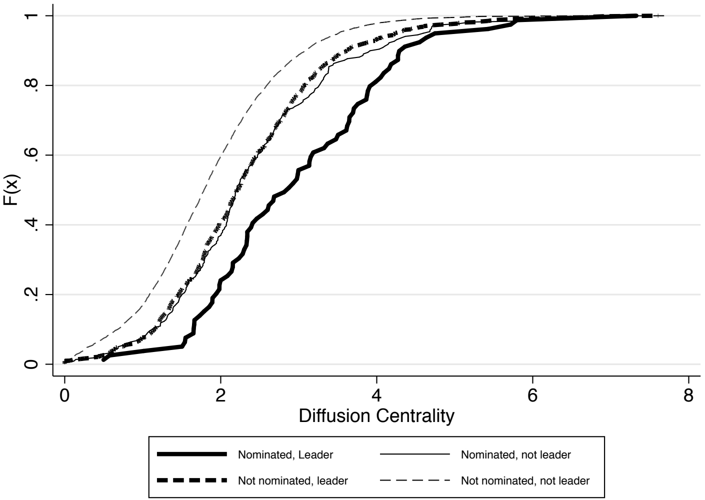
BOTH



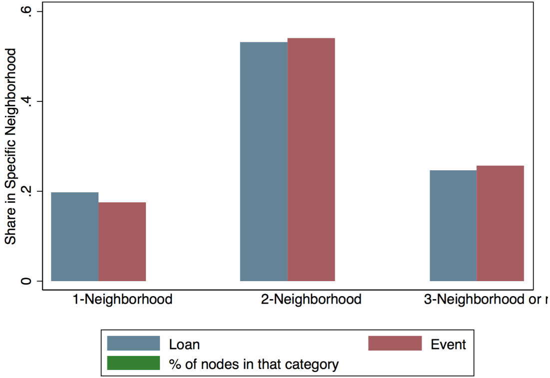
More central, more nomination: Event



More central, more nomination: Loan question



BEYOND IMMEDIATE NEIGHBORHOOD



GUESSES DON'T WORSEN WITH DISTANCE

Num. of nominations regressed on X

Panel A: Poisson Regression	(1)	(2)	(3)	(4)	(5)
Diffusion Centrality	0.607***				
	(0.085)				
Degree Centrality		0.460***			
		(0.078)			
Eigenvector Centrality			0.605***		
			(0.094)		
Leader				0.868***	
				(0.288)	
Geographic Centrality					-0.082
					(0.136)
Observations	6,466	6,466	6,466	5,733	6,466
Panel B: OLS	(1)	(2)	(3)	(4)	(5)
Diffusion Centrality	0.285***				
	(0.060)				
Degree Centrality		0.250***			
		(0.061)			
Eigenvector Centrality			0.283***		
			(0.064)		
Leader				0.422**	
				(0.172)	
Geographic Centrality					-0.025
					(0.038)
Observations	6,466	6,466	6,466	5,733	6,466

NUM. OF NOMINATIONS REGRESSED ON X_1, X_2

Panel A: Poisson Regression	(1)	(2)	(3)	(4)	(5)
Diffusion Centrality	0.642***	0.354**	0.553***	0.606***	0.607***
	(0.127)	(0.176)	(0.098)	(0.085)	(0.085)
Degree Centrality	-0.039				
	(0.101)				
Eigenvector Centrality		0.283			
		(0.186)			
Leader			0.541*		
			(0.305)		
Geographic Centrality				-0.082	
				(0.142)	
Observations	6,466	6,466	5,733	6,466	6,466
Post-LASSO					✓

Panel B: OLS	(1)	(2)	(3)	(4)	(5)
Diffusion Centrality	0.303***	0.161*	0.278***	0.285***	0.285***
	(0.091)	(0.087)	(0.069)	(0.060)	(0.060)
Degree Centrality	-0.020				
	(0.066)				
Eigenvector Centrality		0.138			
		(0.095)			
Leader			0.297		
			(0.175)		
Geographic Centrality				-0.026	
				(0.039)	
Observations	6,466	6,466	5,733	6,466	6,466
Post-LASSO					✓

Understanding gossip nominations

To what extent

- is faster diffusion mediated by only diffusion centrality of gossip seeds?
- does it also reflect the villagers' ability to capture other dimensions of the individual that makes them good at diffusing information?

Is the Gossip effect entirely accounted for by centrality?

VARIABLES	(1) Calls Received	(2) Calls Received	(3) Calls Received	(4) Calls Received
At least 1 Gossip	5.712	7.504	6.047	3.165
	(1.879)	(3.362)	(3.439)	(4.402)
At least 1 Elder	-1.906	-7.853	-7.789	-5.661
	(2.125)	(2.959)	(3.105)	(4.589)
At least 1 High *DC* Seed			4.029	
			(2.619)	
Constant	7.045	8.262	7.282	7.359
	(2.024)	(4.806)	(4.820)	(5.013)
Observations	212	68	68	68
P-value from Joint Test	0.00452	0.0306	0.0525	0.0556
Seed *DC* Control				✓

Is the Gossip effect entirely accounted for by centrality?

VARIABLES	(1) Calls Received Seeds	(2) Calls Received Seeds	(3) Calls Received Seeds	(4) Calls Received Seeds
At least 1 Gossip	1.171	1.516	1.242	0.777
	(0.468)	(0.791)	(0.814)	(1.077)
At least 1 Elder	-0.556	-1.857	-1.846	-1.559
	(0.528)	(0.702)	(0.748)	(1.295)
At least 1 High *DC* Seed			0.758	
			(0.651)	
Constant	2.208	2.480	2.296	2.237
	(0.566)	(1.280)	(1.286)	(1.303)
Observations	212	68	68	68
P-value from Joint Test	0.0166	0.0357	0.0888	0.131
Seed *DC* Control				✓

CONCLUSION

- In practice, network members seem to be good at identifying those who are good at transmitting information
- We highlight one possible mechanism
- However it is clear that it is not the only thing: they do even better than theory.
- Very relevant for policy.