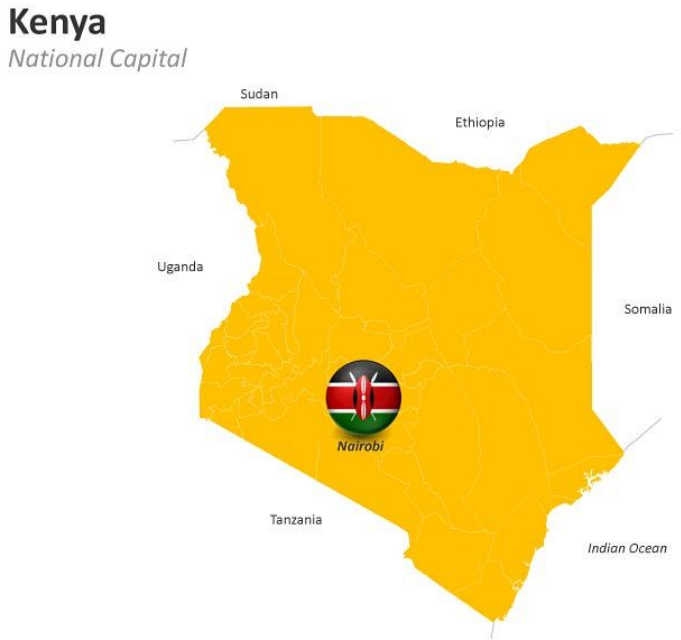
Network Analysis

with an Application in Social Contacts in Kenya

Kristen Li

Background



Datasets:

The study was conducted in the Matsangoni sub-location within the Kilifi Health and Demographic Surveillance Site (KHDSS), coastal Kenya. Five households (L, F, E, B, H) were selected at random from a group of households that had earlier participated in a study to investigate ‘who acquires infection from whom’ (WAIFW).

Contact has two types:

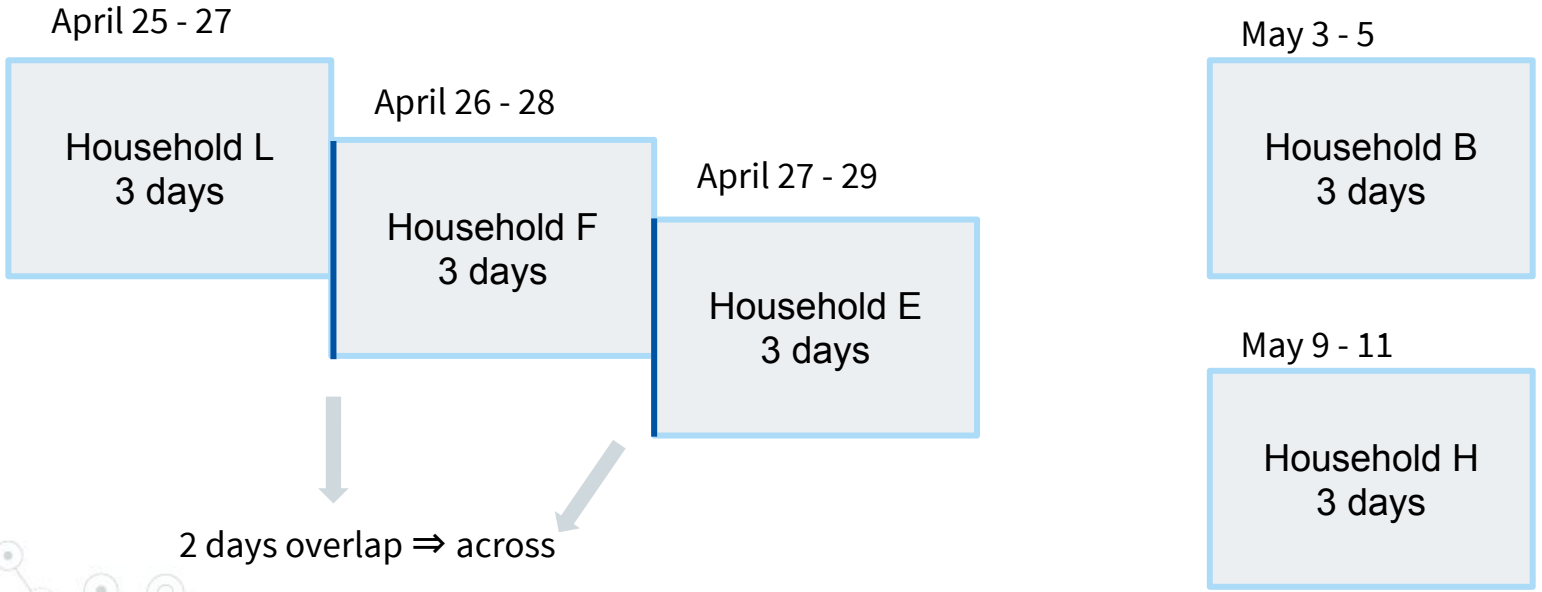
- Contacts between members of the same household (available for all the households);
- Contacts between members of different households (available only for households E, F, and L).

Variables

Variables	Description
h1	household ID of member 1 (L, F, E, B, H)
m1	anonymous ID number of member 1
h2	household ID of member 2 (L, F, E, B, H)
m2	anonymous ID number of member 2
age1	Age 0-5 (0) | Age 6-14 (1) | Age 15-19 (2) | Age 20-49 (3) | Age >=50 (4)
sex1	gender of member 1 (F, M)
age2	Age 0-5 (0) | Age 6-14 (1) | Age 15-19 (2) | Age 20-49 (3) | Age >=50 (4)
sex2	gender of member 2 (F, M)
duration	duration is the duration of the contact event in seconds
day	the day of experiment (1, 2, 3)
hour	the day time of the contact event (7 – 20)

Variables: frequency

Within household frequency = count/3
Across household frequency = count/2



Descriptive Statistics

Household	Total number of participants	Female vs. Male	Age 0-5	Age 6-14	Age 15-19	Age 20-49	Age >=50
B	15	12:3	4	4	1	4	2
E	17	8:9	3	7	0	7	0
F	8	4:4	1	1	2	3	1
H	29	22:7	6	13	2	6	2
L	6	4:2	1	1	1	3	0

- Contacts between members of the same household: 75 members, 32425 observations;
- Contacts between members of different households: 28 members (only 3 out of 6 in L have across household contacts), 218 observations.

Research Question

1. **Who are the more powerful individuals in the social network?**

2. **Who are more likely to contact with others?**
   - **Is there any different contact features between within household contacts and across household contacts?**

3. **What are the characteristics of the social network?**

Methodology - Nodes (Individuals)

Degree Centrality: the degree a node represents the number of distinct individuals with whom an individual has been in contact during the time window
⇒ How many direct, ‘one hop’ connections each node has to other nodes within the network.

Betweenness Centrality: quantifies the number of times a node acts as a bridge along the shortest path between two other nodes.
⇒ Who’s more likely to be a bridge when spread the infection?

Eigen Centrality: measures a node’s influence based on the number of links it has to other nodes within the network and also take into account how well connected a node is, and how many links their connections have, and so on through the network.
⇒ Who’s the central node with central neighbors?

Methodology - Edges (Undirected)

Frequency: the weight an edge between nodes is the number of contact events each day on average recorded between these individuals during the time window;

Duration: the weight an edge between nodes is the average of duration of the contacts recorded during the time window between the two individuals.

1. Within households

Descriptive Statistics

Age	# participants	# contacts
0(0-5)	15	10804
1(6-14)	26	15462
2(15-19)	6	607
3(20-49)	23	4835
4(>=50)	5	717
ALL	75	32425

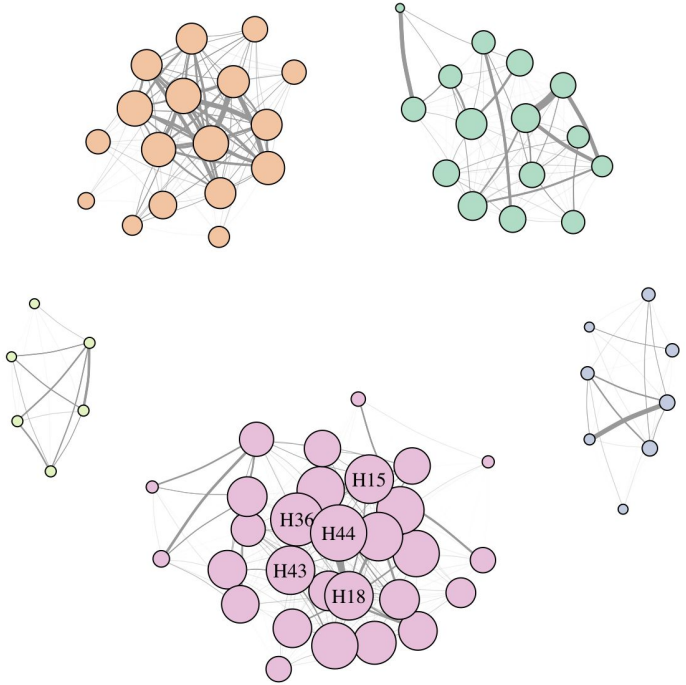
144 contacts per person per day

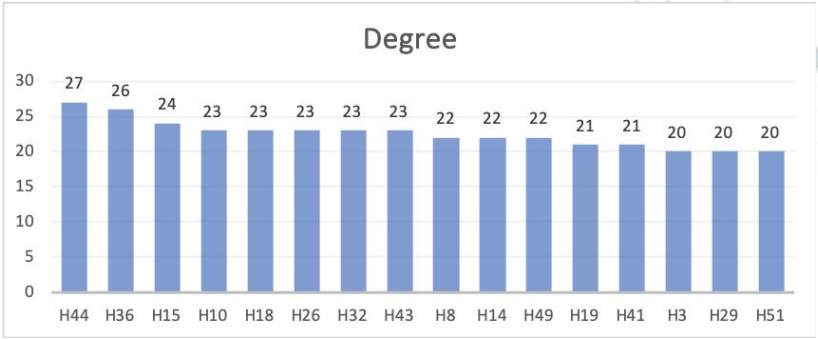
Within household:

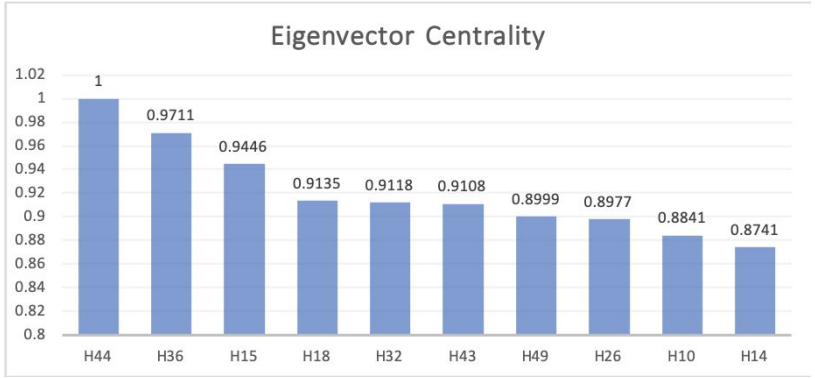
network of contact frequency within households color coded **by households**

Node size: **degree centrality**, edge width: frequency, labeled high eigenvalue members

Powerful members: H44, H36, H18, H15



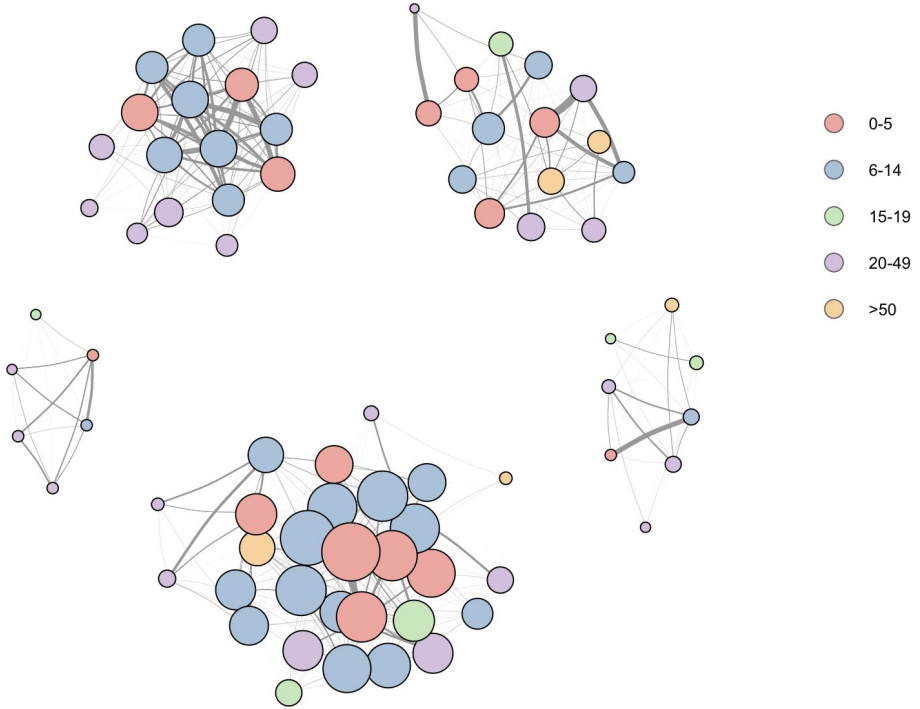




Within household:

network of contact frequency within households color coded by **age**

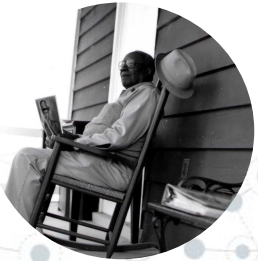
Node size: **degree centrality,** edge width: frequency,
Powerful members: H44, H36, H18, H15, H43



In each household:

Children (Age 0-5 and age 6-14) are frequently connected together.

Parents and kids (Age 20-29 and Age 0-5 or Age 20-29 and Age 6-14) are frequently connected.



Older adults (age >=50) barely connect to others.

Within household:
network of contact numbers within households by **age**

Total number of contacts group by age

	Age 0-5	Age 6-14	Age 15-19	Age 20-49	Age >=50
Age 0-5	3070	5587	56	1847	244
Age 6-14	5331	7802	161	2023	145
Age 15-19	57	170	79	241	60
Age 20-49	1787	1975	226	592	255
Age >=50	199	128	70	245	75

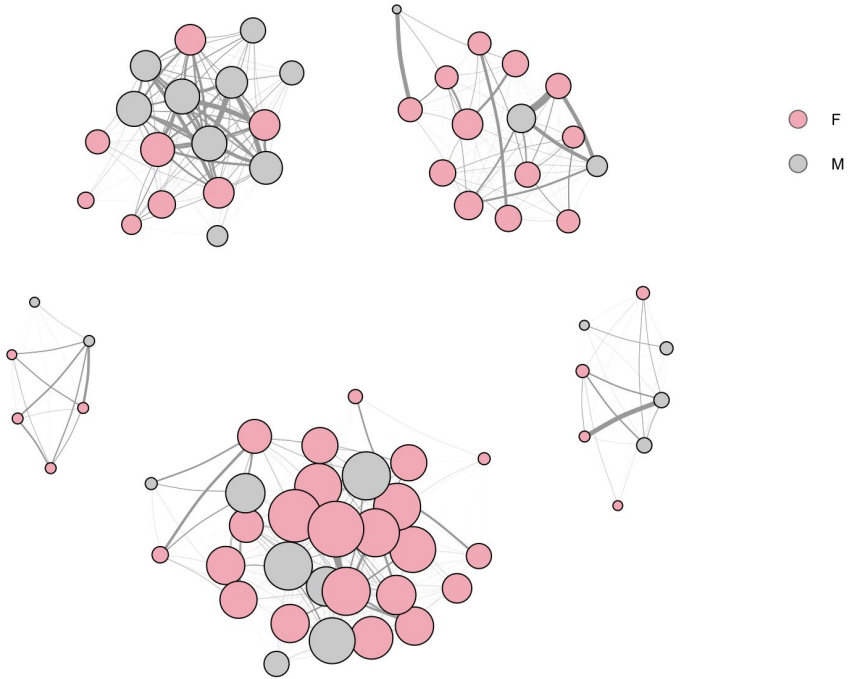
Average number of contacts per person group by age

	Age 0-5	Age 6-14	Age 15-19	Age 20-49	Age >=50
Age 0-5	205	372	4	123	16
Age 6-14	205	300	6	78	6
Age 15-19	10	28	13	40	10
Age 20-49	78	86	10	26	11
Age >=50	40	26	14	49	15

Within household:

network of contact frequency between households color coded by **sex**

Same sex contact is more often



Within each household:

- Same sex contacts are more frequently.

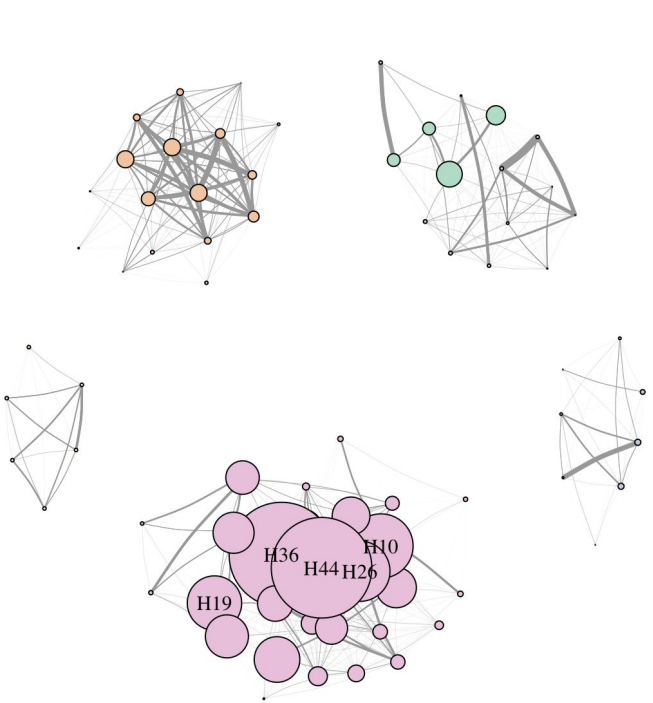
Number of contacts: more same gender contacts

	F(#)(50)	M(#)(25)	F(Avg)	M(Avg)
F(50)	11084	7408	222	148
M(25)	6842	7091	274	284

Within household:

network of contact frequency between households color coded by households

Node size: **betweenness centrality**, edge width: frequency



Members with high betweenness centrality and eigenvalues

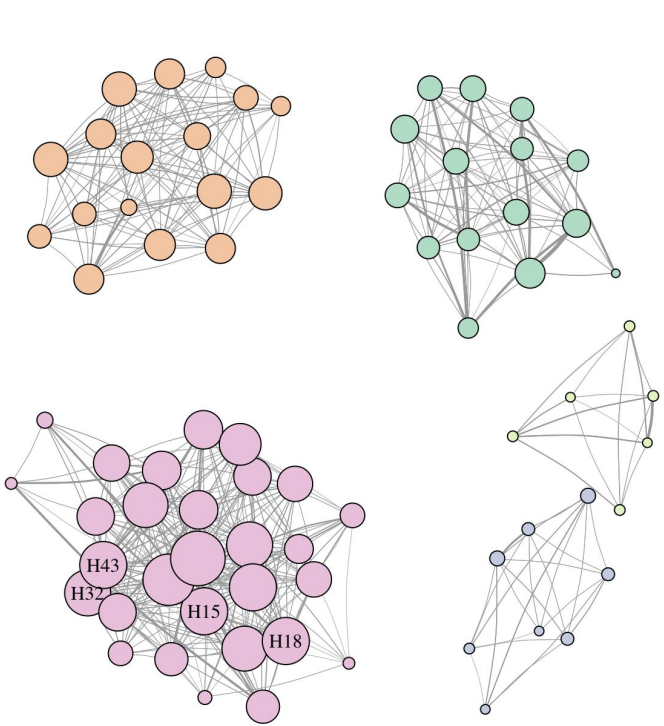
Member	H36	H44	H19	H10	H26	H15
Degree	17.96	16.22	9.10	7.92	7.30	6.81
Member	H43	H15	H32	H18	H29	
Eigenvalue	1.0000	0.9122	0.9050	0.9038	0.8883	

Within household:

network of average contact duration between households color coded by households

Node size: **degree centrality**, edge width: avg duration

Powerful members: H44, H43, H32, H15, H18



Members with high degree scores and eigenvalues

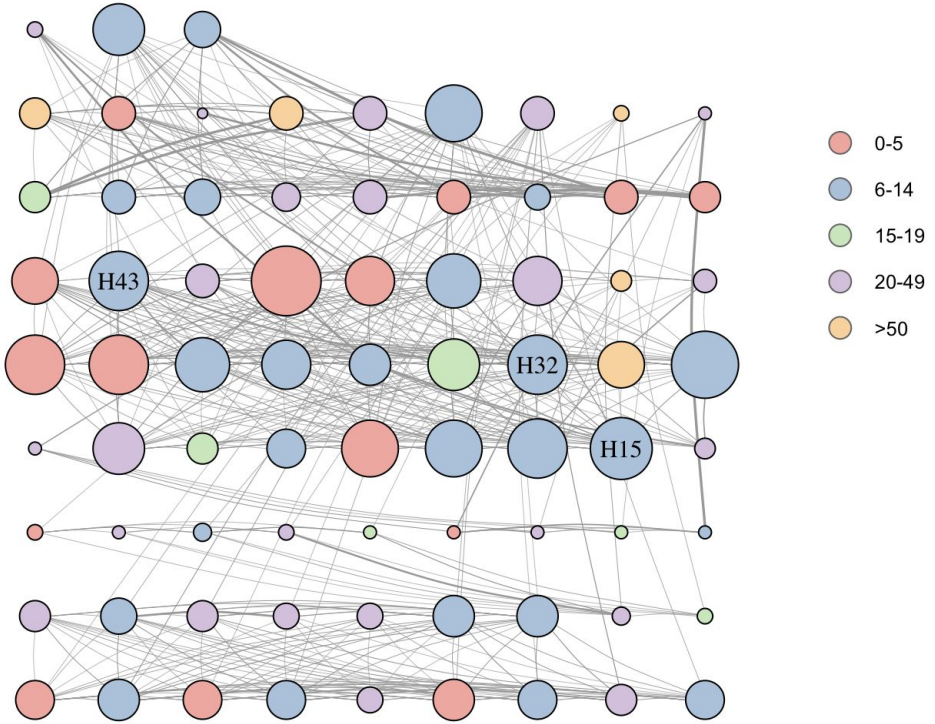
Member	H44	H36	H15	H18	H26	H32
Degree	27	26	24	23	23	23
Member	H43	H15	H32	H18	H29	H44
Eigenvalue	1.00	0.92	0.91	0.89	0.89	0.87

Within household:

network of average contact duration between households color coded by age

Node size: **degree centrality**, edge width: avg duration

Powerful members: H43, H32, H15



Within households:

- Children (Age 0-5 and age 6-14) are connected together for longer duration.
- Parents and kids (Age 20-29 and Age 0-5 or Age 20-29 and Age 6-14) are connected longer.

Within household:

network of contact duration within households by **age**

Sum of contact duration group by age

	Age 0-5	Age 6-14	Age 15-19	Age 20-49	Age >=50
Age 0-5	92280	191960	1600	101280	7280
Age 6-14	182520	244840	6960	97980	3300
Age 15-19	1600	6380	2060	19820	5400
Age 20-49	97460	91180	20460	26220	8540
Age >=50	5740	3180	5380	7980	3440

Average contact duration for each person group by age

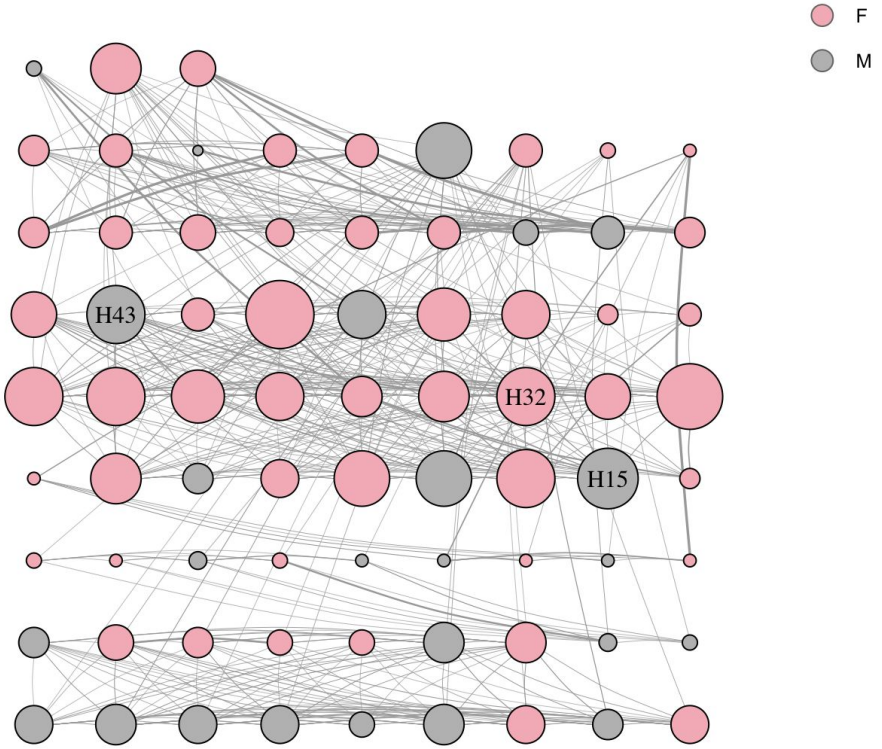
	Age 0-5	Age 6-14	Age 15-19	Age 20-49	Age >=50
Age 0-5	6152	12797	107	6752	485
Age 6-14	7020	9417	268	3768	127
Age 15-19	267	1063	343	3303	900
Age 20-49	4237	3964	890	1140	371
Age >=50	1148	636	1076	1596	688

Within household:

network of average contact duration between households color coded by **sex**

Node size: **degree centrality**, edge width: avg duration

Powerful members: H43, H32, H15



Within each household:

- Same sex contacts have slightly longer duration.

Duration of contacts: more same gender contacts

	F(dur)	M(dur)	F(Avg dur)	M(Avg dur)
F(50)	403480	309140	8070	6183
M(25)	288360	233860	11534	9354

2. Across households

Descriptive Statistics

Age	# participants
0(0-5)	15
1(6-14)	26
2(15-19)	6
3(20-49)	23
4(>=50)	5
ALL	75

61 edges in total

144 contacts per person per day

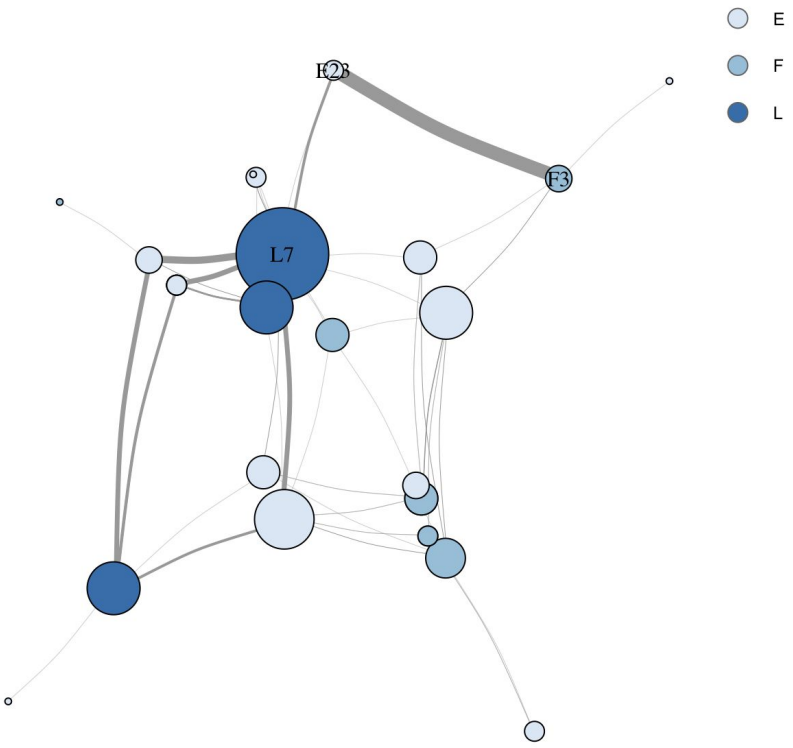
3 members of household L did not record contacts with members of other households

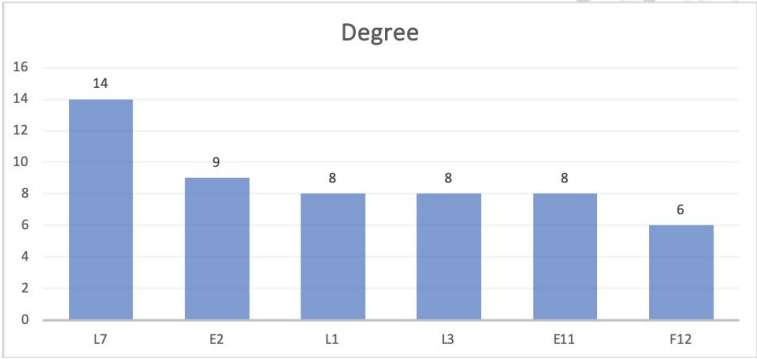
Across household:

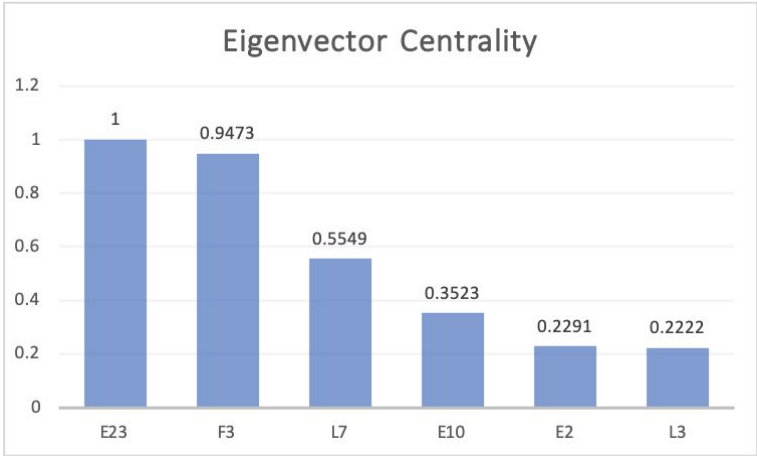
network of contact frequency across households color coded by **households**

Node size: **degree centrality**, edge width: frequency

Powerful members - eigenvalue: L7, E23, F3





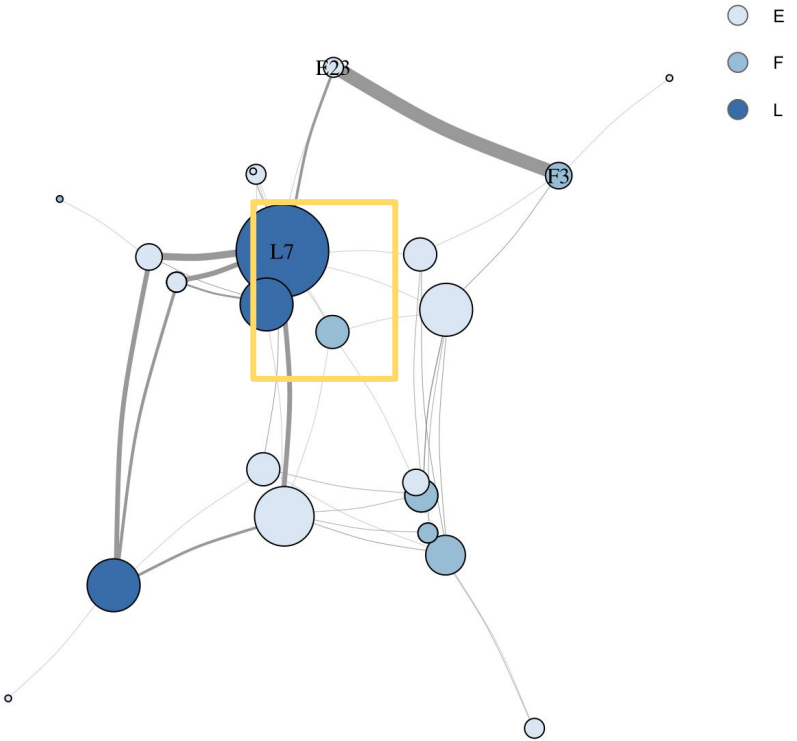


Across household:

network of contact frequency across households color coded by **households**

Node size: **degree centrality**, edge width: frequency

Powerful members - eigenvalue: L7, E23, F3



- Household E and F, E and L are connected closely.
- However, household F and L are connected with 1 edge only (L7 and F11, both female and 20-49 years old) and only happened once.

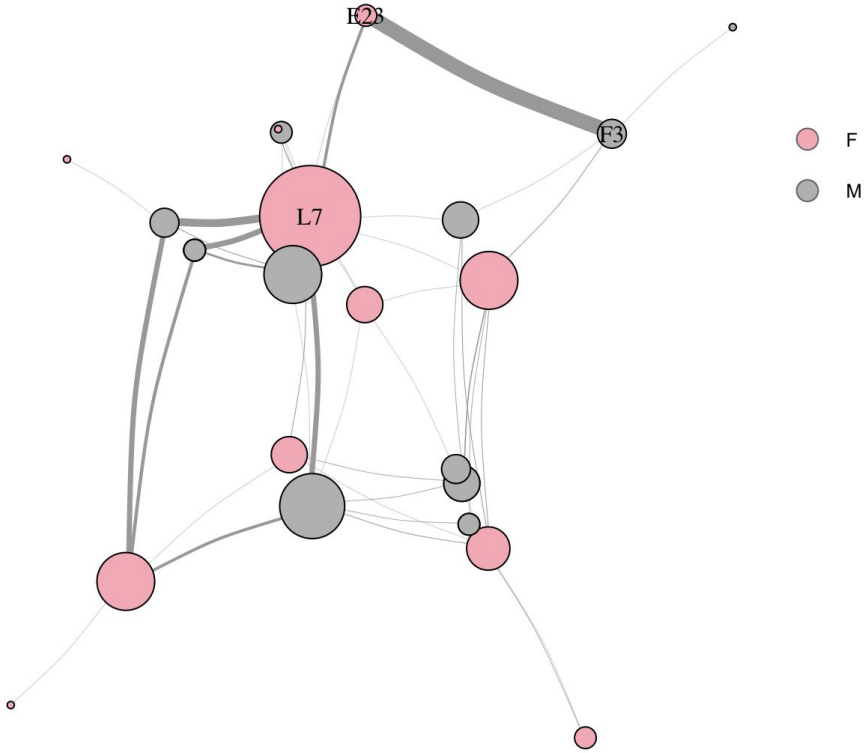
⇒ household F and L are not closely connected.

⇒ 2 households are connected because 1 pair of individuals.

Across household:

network of contact frequency across households color coded by **sex**

Node size: **degree centrality**, edge width: frequency
Powerful members - eigenvalue: L7, E23, F3



Across households:

- Across household contacts are more likely to happen between different genders.

Friendships, relationships, colleagues, etc

Across household:
network of contact across households by **sex**

Total number of contacts group by age

	F(14)	M(14)
F	54	86
M	58	20

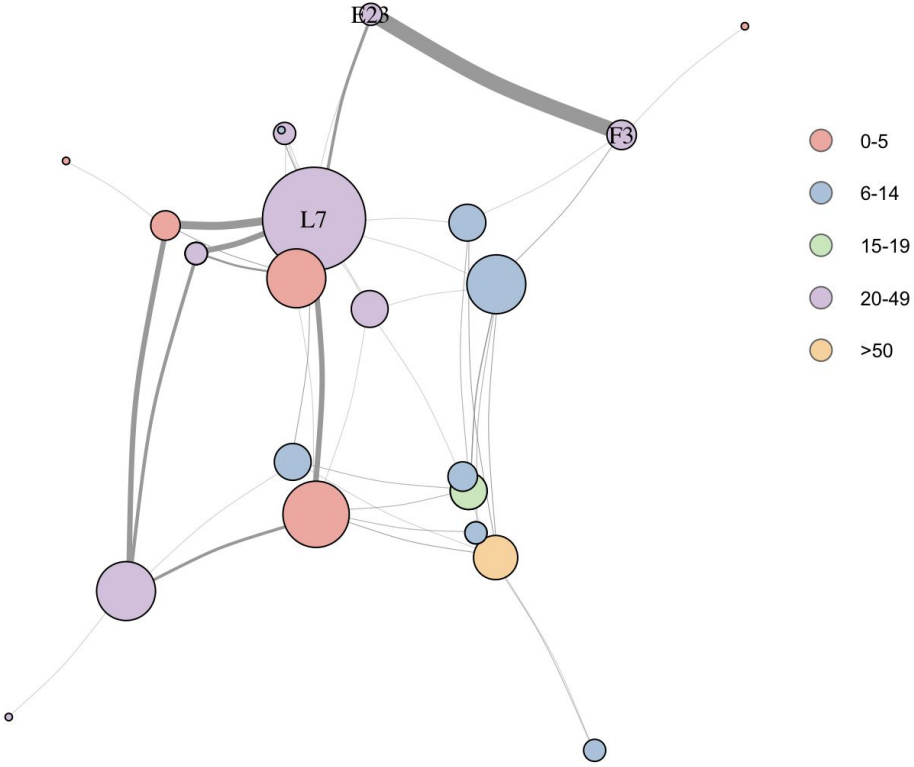
Average number of contacts per person group by age

	F(14)	M(14)
F	4	6
M	4	1

Across household:

network of contact frequency across households color coded by **age**

Node size: **degree centrality**, edge width: frequency
Powerful members - eigenvalue: L7, E23, F3



- Contacts across households are also connected by people from different age group ,mostly between adults and children.
- Age 20-49 from different households have frequent contacts.

Across household:
network of contact across households by **age**

Total number of contacts group by age

	Age 0-5 (5)	Age 6-14 (8)	Age 15-19 (2)	Age 20-49 (12)	Age >=50 (1)
Age 0-5	4	1	2	19	2
Age 6-14		2	11	8	7
Age 15-19					
Age 20-49	56	7		99	
Age >=50					

Average number of contacts per person group by age

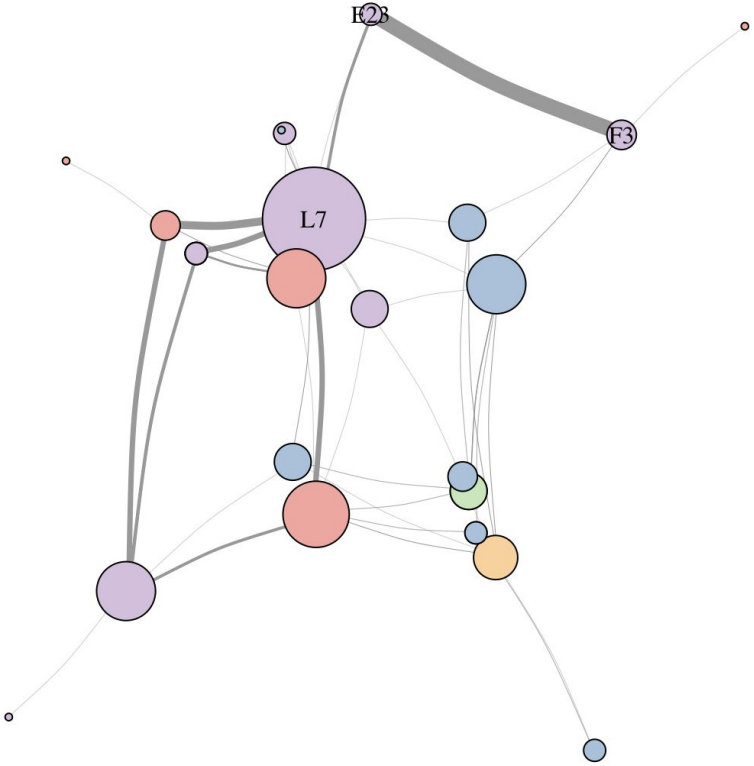
	Age 0-5 (5)	Age 6-14 (8)	Age 15-19 (2)	Age 20-49 (12)	Age >=50 (1)
Age 0-5	1	0	0	4	0
Age 6-14		0	1	1	1
Age 15-19					
Age 20-49	5	1		8	
Age >=50					

Across household:

network of contact frequency across households color coded by **age**

Node size: **degree centrality**, edge width: frequency

Powerful members - eigenvalue: L7, E23, F3



Babysitters and kids play in a common area.

Parents attend the same parental group.

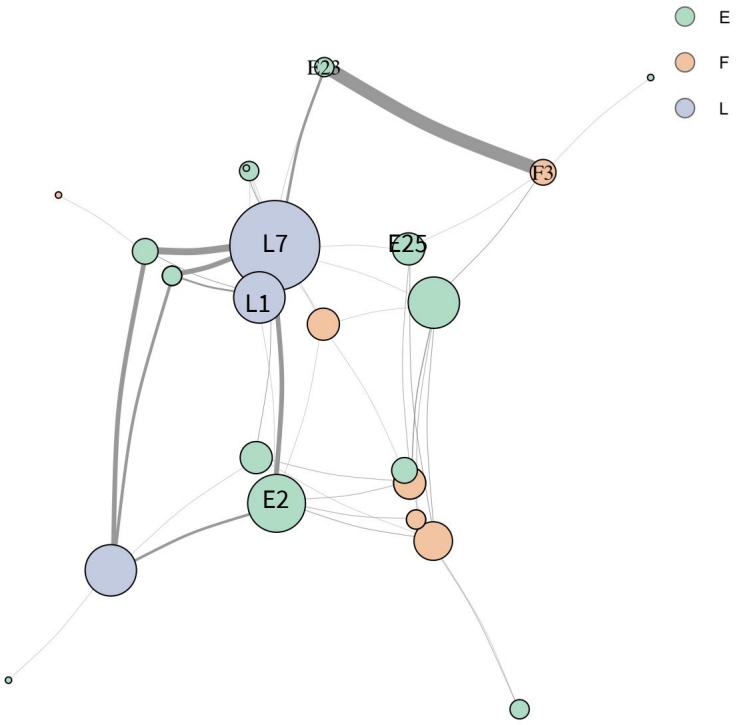
Friends, colleagues, couples, etc

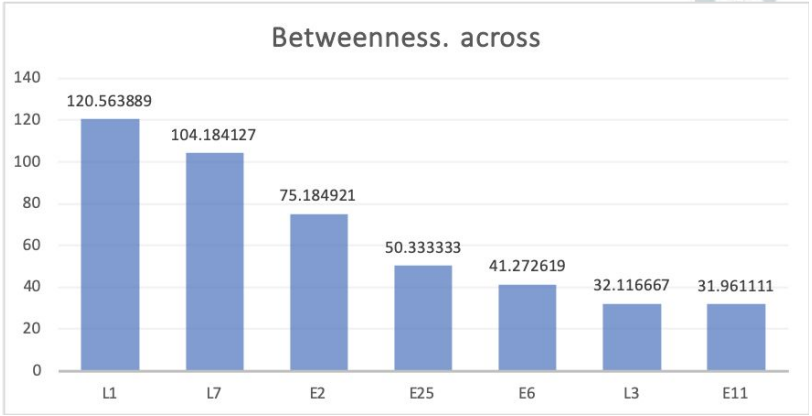
Across household:

network of contact frequency across households color coded by **households**

Node size: **betweenness centrality**, edge width: frequency

Powerful members: L7, L7, E2, E25





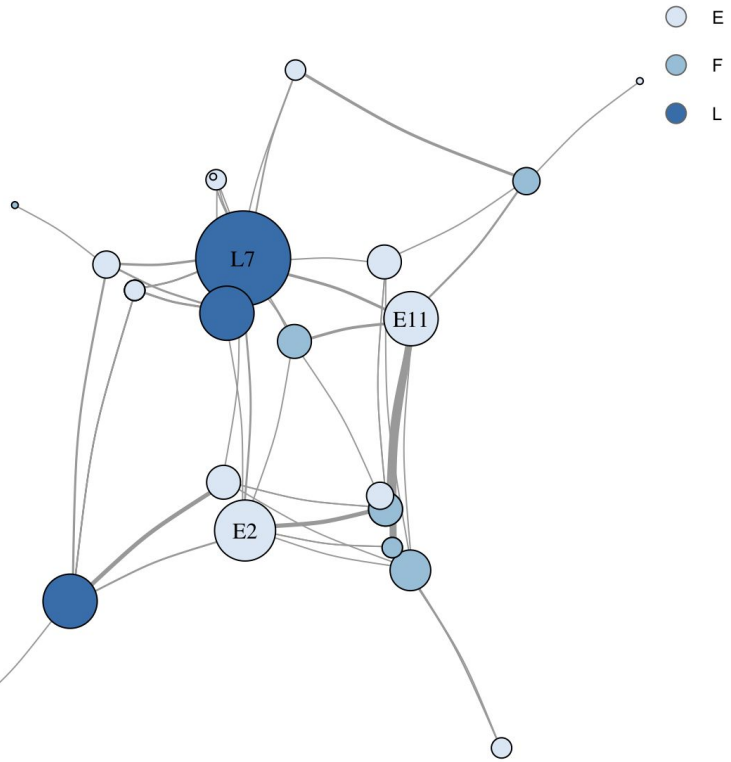
Best bridge: L7, L7, E2, E25

Across household:

network of average contact duration across households color coded by **households**

Node size: **degree centrality**, edge width: avg duration

Powerful members: L7, E2, E11



- Contacts across households are connected longer by people between adults and children.
- Age 20-49 from different households have longer time contacts.
- E11 and F10 only contact once but stayed together for quite a long time.
- Interestingly, the couple met quite frequently but doesn't stay together for a long time.

Across household:

network of average contact duration across households color coded by **households**

Node size: **degree centrality**, edge width: avg duration

Powerful members: L7, E2, E11

Sum of duration of contacts group by age

	Age 0-5 (5)	Age 6-14 (8)	Age 15-19 (2)	Age 20-49 (12)	Age >=50 (1)
Age 0-5	100	20	100	460	40
Age 6-14		120	260	200	140
Age 15-19					
Age 20-49	1780	200		1980	
Age >=50					

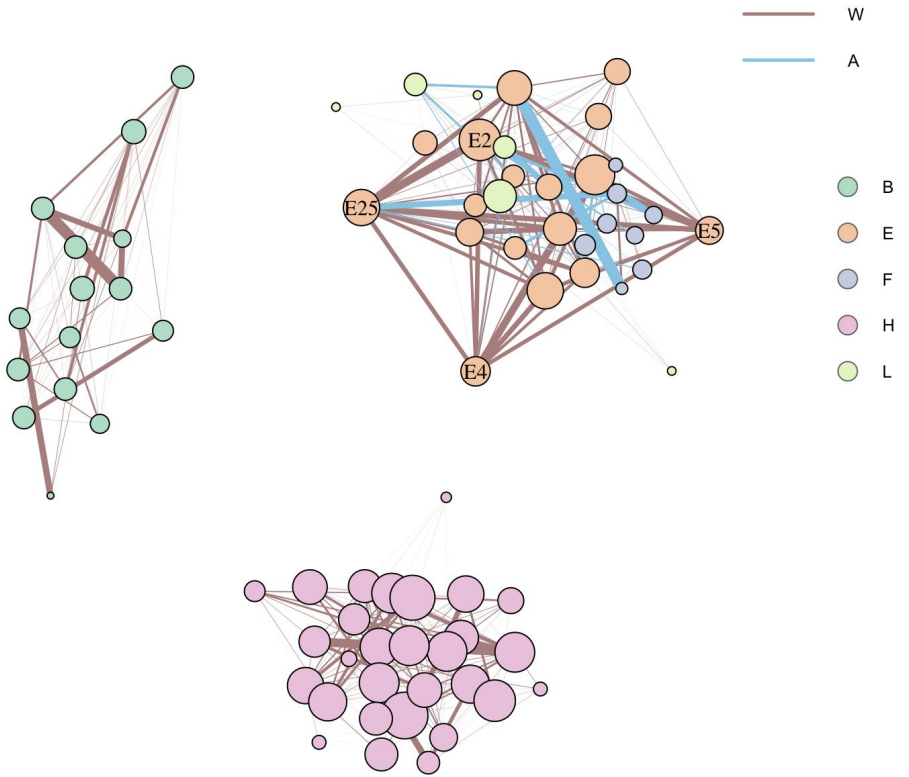
Average duration of contacts per person group by age

	Age 0-5 (5)	Age 6-14 (8)	Age 15-19 (2)	Age 20-49 (12)	Age >=50 (1)
Age 0-5	20	4	20	92	8
Age 6-14		15	33	25	18
Age 15-19					
Age 20-49	148	17		165	
Age >=50					

3. Full contact network

Entire Contact Network: both within and across households

Node: degree, edge = frequency
powerful individuals - eigenvalue: E11, E27, E10, E25, E4, E30, E20



Nodes:

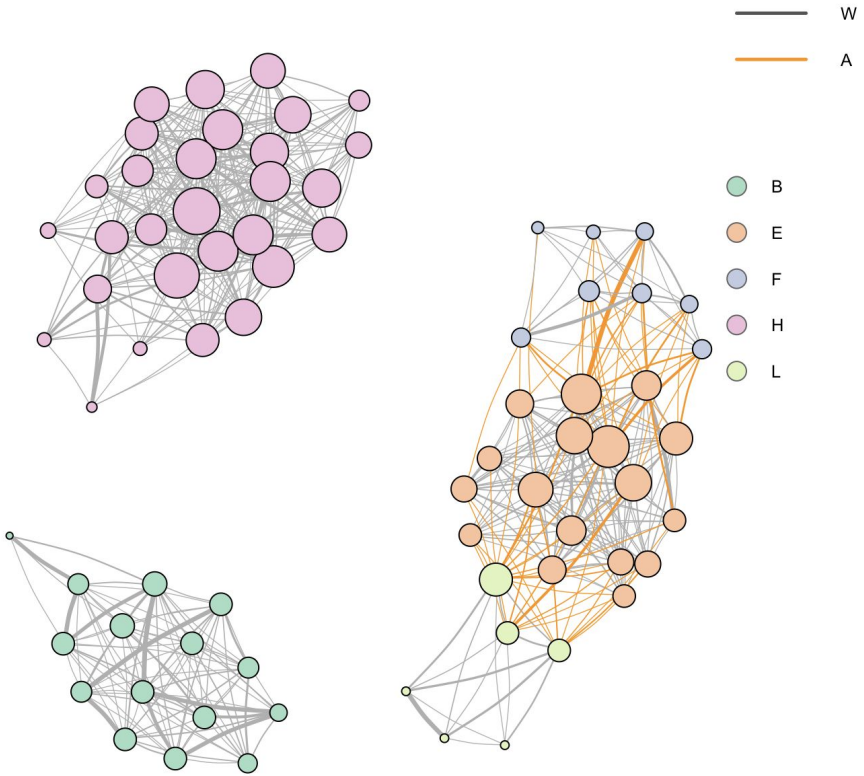
75 nodes, proportional to its degree

Edges:

576 edges, proportional to the weight of frequency per day

Entire Contact Network: both within and across households

Node: degree, edge = avg duration
powerful individuals - eigenvalue: E25, E4, E2, E5



Nodes:

75 nodes, proportional to its degree.

Edges:

576 edges, proportional to the weight of average time spent between 2 individuals.

Conclusion

Within household, same sex contacts are more frequently and stay longer;
Across households, different sex contacts re more often.

Within household, contacts between kids (0-19) and kids (0-19), between adults (20-49) and kids (0-19) are more frequently and stays longer.
Across households, contacts between adults (20-49) and kids (0-19) are more frequently and stays longer.

Limitations

- Across household contacts are very unstable and more often, diseases are transmitted from one household to another by individual contacts.
  ⇒ longer period of across household contact recording is necessary.
  ⇒ more across households samples.

- The 5 samples are not representative or large enough to study the relationship between diseases spread and social networks.
  ⇒ larger sample size