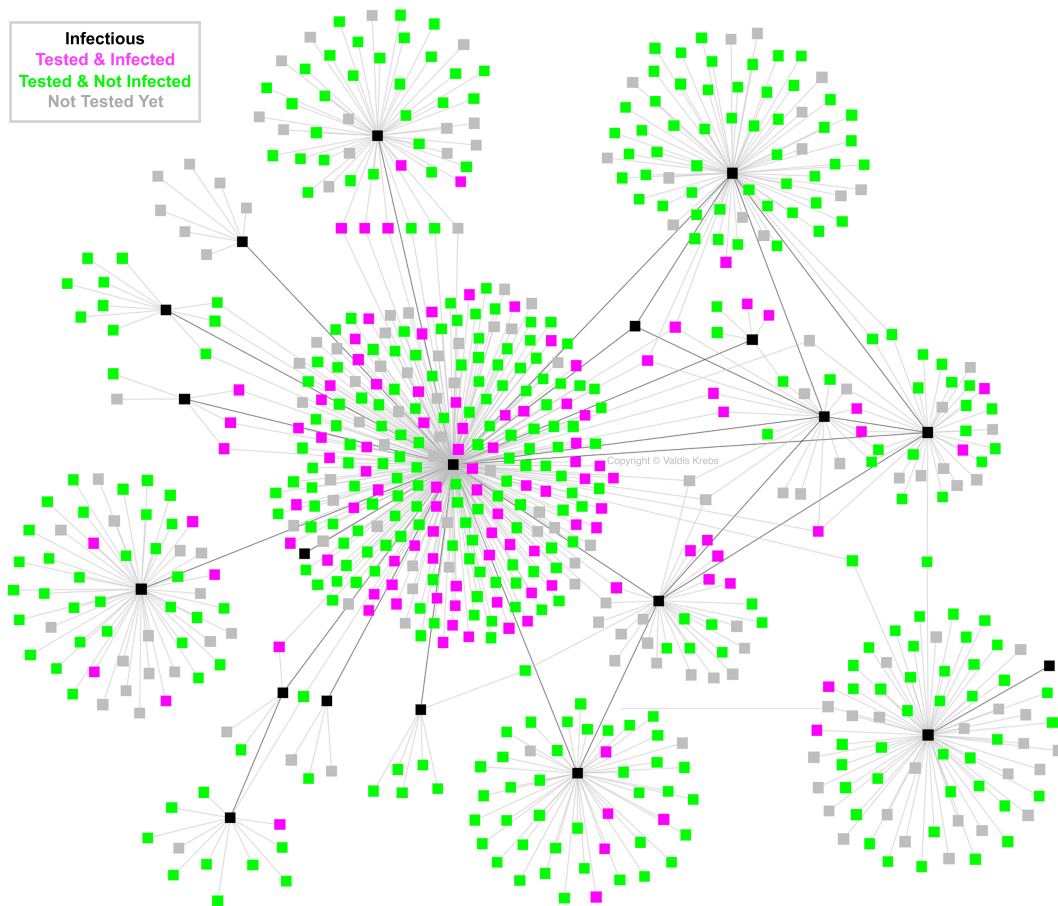




# Transmission Network Analysis

tracking and stopping the spread of contagious diseases



**Valdis Krebs**

Chief Scientist

Orgnet, LLC

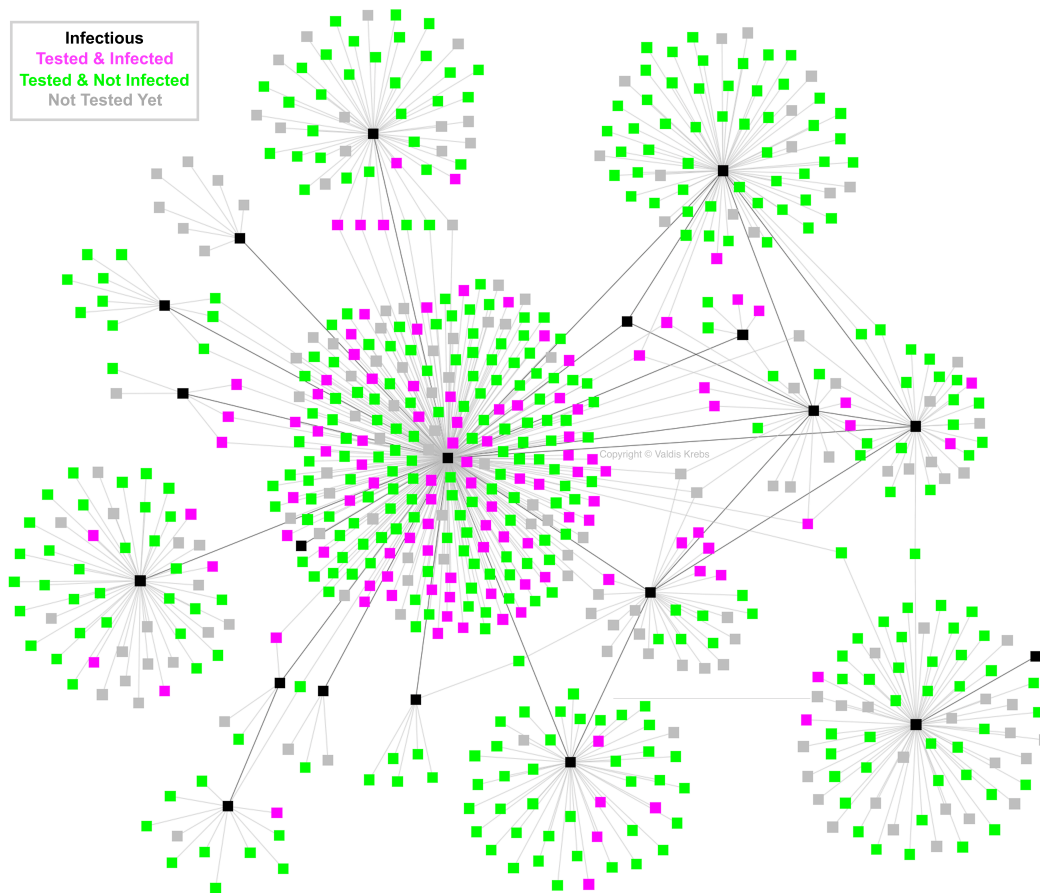
[valdis@orgnet.com](mailto:valdis@orgnet.com)

Both good and bad contagions spread via human networks. For good contagions, like ideas, knowledge, and experience we want our networks to quickly and completely spread the information. For bad contagions, like disease or disinformation, we want to quickly spot the source(s) and disrupt the spread.

We currently have an outbreak of **Covid-19**, a.k.a. **SARS-Cov-2** or **Coronavirus**. It is spreading around the world through human contact. Like many infectious diseases, Covid-19 spreads through airborne spray from coughs and sneezes of infected persons. It can also spread from the infectious touching various surfaces after they have sneezed or coughed into their hand.



Below, in Figure 1, is a map of an actual disease outbreak. It is not Covid-19, but a disease that also spreads like Covid-19 - via airborne particles or surfaces in the vicinity of infectious persons. The map below is a TB (tuberculosis) outbreak that happened in southwest America over 10 years ago.



**Figure 1**

Each node/square represents a person, and each link/line represents a contact between two people. A black node represents *infectious* persons, a magenta/pink node represents *infected* people, and a green node shows those exposed but not infected. Both the pink and green nodes' status was determined *after testing* them for the disease. Gray nodes were recorded as contacts of the black infectious nodes, but have not been tested as of time of this snapshot. The darker lines (links) represent ties/links between the infectious people, and the

light gray lines (links) show the ties/links between the infectious and their contacts.

## Contact Tracing

Epidemiologists do *contact tracing* in order to track and stop a disease outbreak. First, they find the sick and treat them while isolating them to prevent further transmission. Next, they find who the sick have recently had contact with. **All contacts need testing and monitoring!** The battle line against the spread of a contagious disease is drawn around each infectious person and their contacts. If the contacts are tested in time, and the ones showing symptoms are isolated, the continued spread around this person should stop there.

Knowing the contacts of the infected, points public health departments to who requires testing and then who requires monitoring and/or treatment. Each newly discovered infectious case requires a new round of contact tracing around them - who have they had contact with while they were infectious (before and after they were showing obvious symptoms).

A contact network is not necessarily a social network. Network contacts might be family, friends, acquaintances, or strangers. We can pick up a disease from a sick family member at home or from a stranger via an inopportune sneeze in a crowded coffee shop. Human networks evolve into what social scientists call "small world networks" - we tend to cluster together via the social dynamic of homophily (i.e. birds of a feather flock together). A small-world network is made up of connected clusters where there are more connections *within* the cluster than *between* the clusters. This creates many redundant ties in our closest networks, which are very effective at spreading whatever is flowing in those networks. Families that live together have a link between everyone in the dwelling. Most close friends of a person are also friends with each other, and colleagues at

work usually also mostly work with each other. We end up in clustered communities - good for spreading ideas, knowledge and affection, but also good for spreading disease. We tend to spend more time with family, friends and colleagues, so this gives more time and opportunity for any virus to spread. The Centers for Disease Control (CDC) distinguishes between *close contacts* and *casual contacts* when tracing an infectious person's contacts.

But contact networks are not just made up of people we know; they also include general acquaintances (the mail person, the barista at the coffee shop) and total strangers (the traveler at the airport, or the person sitting next to you on the bus). We have both *known* and *unknown* contacts. The connections with our known contacts spread the disease within our clusters/communities. But the connections with unknown contacts can be worse (for the outbreak) - they can spread the disease *between* clusters that were previously separated. A bad choice of seats on the bus/plane can result in the spread of a disease from the Smith family to the Jones family - who previously had no interaction. The Jones family will never know where their illness came from, but they may know who brought it home.

Below, in Figure 2, we see a common 2 x 2 matrix of how our contact networks are might be composed. We look at two key factors, *Familiarity* (how well do we know the person) and *Proximity* (how close is the contact with the person). There are four combinations of Familiarity and Proximity that describe how each of us comes in contact with the rest of the world.

		<b>FAMILIARITY</b>	
		<b>Known</b>	<b>Unknown</b>
<b>PROXIMITY</b>	<b>Close</b>	Family, Friends, Colleagues you have F2F contact with regularly <i>(those you share dwelling, work and table space with)</i>	Strangers you have F2F contact with regularly <i>(taxi/bus driver, fellow shopper, delivery person, retail clerk, etc.)</i>
	<b>Distant</b>	Family, Friends, Colleagues who live/work in distant places from you	Strangers in distant locations who you have no contact with or awareness of

**Figure 2**

In Figure 3, we have eliminated two boxes in the matrix that we really do not need to worry about with the spread of a disease around us. Those far away cannot infect us, even if we choose to contact them via some electronic media. The key is face-to-face (F2F) interaction – close contact. Both friends and strangers can pass the virus if they are sick and close by.

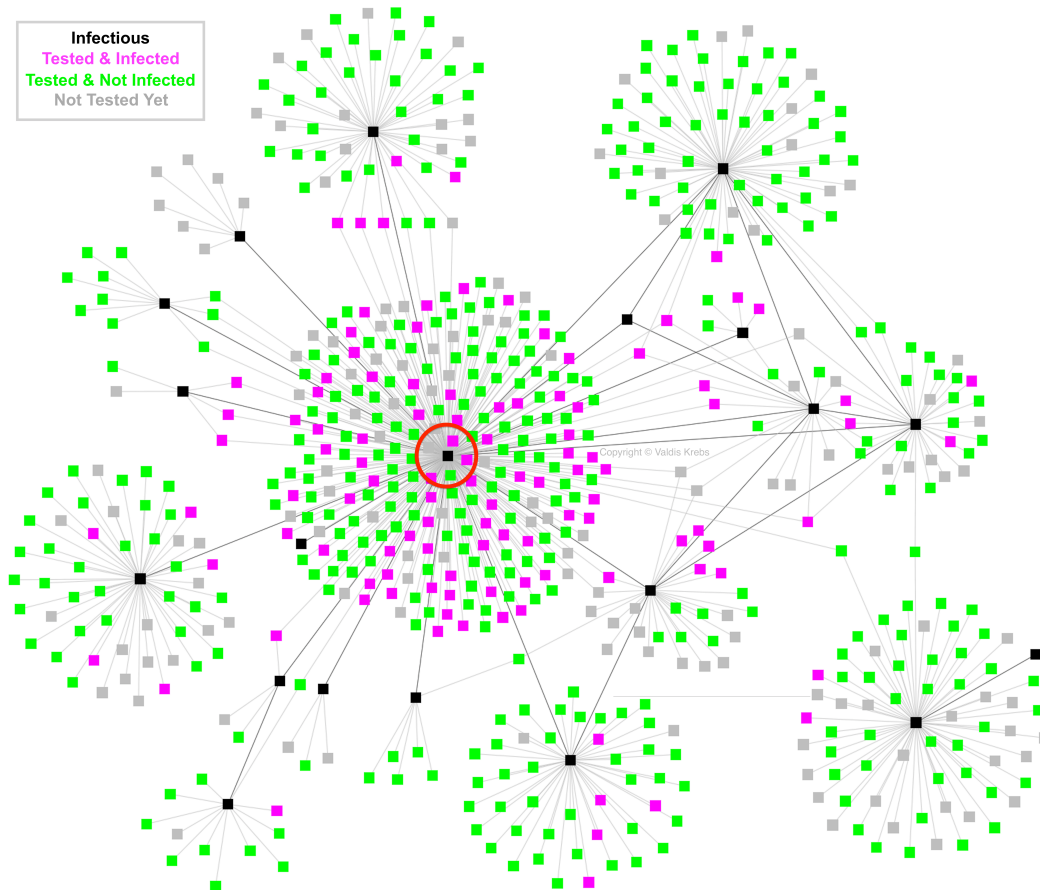
		<b>FAMILIARITY</b>	
		<b>Known</b>	<b>Unknown</b>
<b>PROXIMITY</b>	<b>Close</b>	Family, Friends, Colleagues you have F2F contact with regularly <i>(those you share dwelling, work and table space with)</i>	Strangers you have F2F contact with regularly <i>(taxi/bus driver, fellow shopper, delivery person, retail clerk, etc.)</i>
	<b>Distant</b>	Family, Friends, Colleagues who live/work in distant places from you	Strangers in distant locations who you have no contact with or awareness of

**Figure 3**

## Network Analysis & Visualization

Network visualization was an important step forward of utilizing the lists of contact data already collected in outbreak investigations. Network visualization summarized on one page, the data from many pages of contact tracing data. Attributes of various persons in the outbreak could be quickly designated by node color, shape, size and location. Link direction, thickness and color revealed various attributes of the relationship / flow between the infectious and their contacts. For this demonstration, we keep things simple by just showing the color of both nodes and links.

Looking at the network map in Figure 4 below, our eye naturally gravitates toward the most connected black node in the left-center of the diagram - see the red circle. We look at this person and see he had the most contacts and that he infected the most others (magenta / pink nodes). We might naturally jump to the conclusion: this is a "super-spreader". We assume super-spreaders have some magical power to infect many people. Yet, we usually find that the context of the spread had more to do with infections than some mighty power to infect. We see that this infectious node did infect many, but we also see that he did not infect many others (green nodes).



**Figure 4**

What was so unique about this person? It was not his personal power to infect - it was his longevity. He came into contact with many persons while he was sick, as was not discovered until the outbreak investigation began. He was not only in this city for a long time, he was hard to find. It was determined that he resided in several places during his infectious period, including the county jail where he infected unknown contacts and spread the disease to other social circles.

He was determined to be the *index patient* of this particular outbreak. Once the outbreak investigation began, we can see that none of other infectious persons had anywhere as large of a contact circle as the index patient. The other infectious people were processed via contact tracing in a timely manner and therefore had fewer contacts during

their infectious period – fewer contacts and fewer infections transferred. This begins to show the power of tracing, mapping, and testing. Knowing who is sick, who are the contacts, and who needs to be tested and monitored, is the sine qua non of epidemiologists stopping outbreaks!

Since testing is so important to managing and stopping an outbreak, we need to make sure everyone (if possible) gets tested. Public health workers need to prioritize who will get tested first, and who can wait. How do they do this? Much of it is based on probability - who is likely to get infected? Here we revert to the two contact categories - close and casual. Previous outbreaks have shown that close contacts tend to spread contagions better than casual contacts. More time together, with more possibility of touch and spray from coughing sneezing individuals, lead to higher probabilities of infection. Also individuals exposed to more than one infectious person will have a high probability of becoming infected.

In Figure 5 we see several individuals (gray nodes) who have not been tested and have had contact with two or more infectious individuals (black nodes). These individuals (grey nodes) should be prioritized for testing - they are shown within the red ovals. We see quite a few individuals with connections to two or more infectious individuals (black nodes) have already been uncovered as infected (magenta/pink nodes). These priority-untested individuals can also be discovered using social network analysis metrics that analyze a node's connectivity and *reach* in a network.

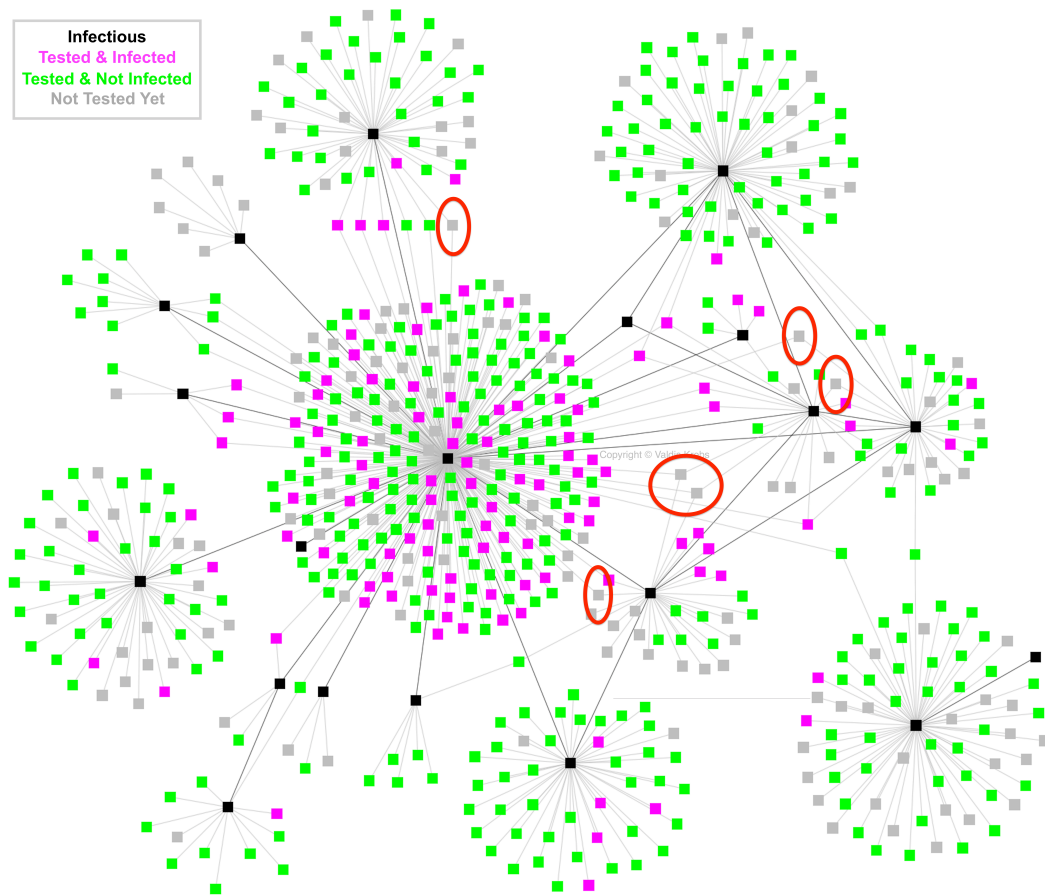


Figure 5



What else does our network diagram of the outbreak reveal to us? Figure 6 below shows how many social circles were involved in this outbreak. The red oval shows the family / friends of the index patient - these were his known contacts. The other infected individuals (black nodes) are not connected to each other and therefore probably do not know each other and are members of different social circles - these were the unknown contacts of the index patient, people he probably interacted with in public places that he frequented. It is these infected unknown contacts that take an infection to parts of the network where it is not already active. They are the bridges between the small-worlds (clusters) in human networks.

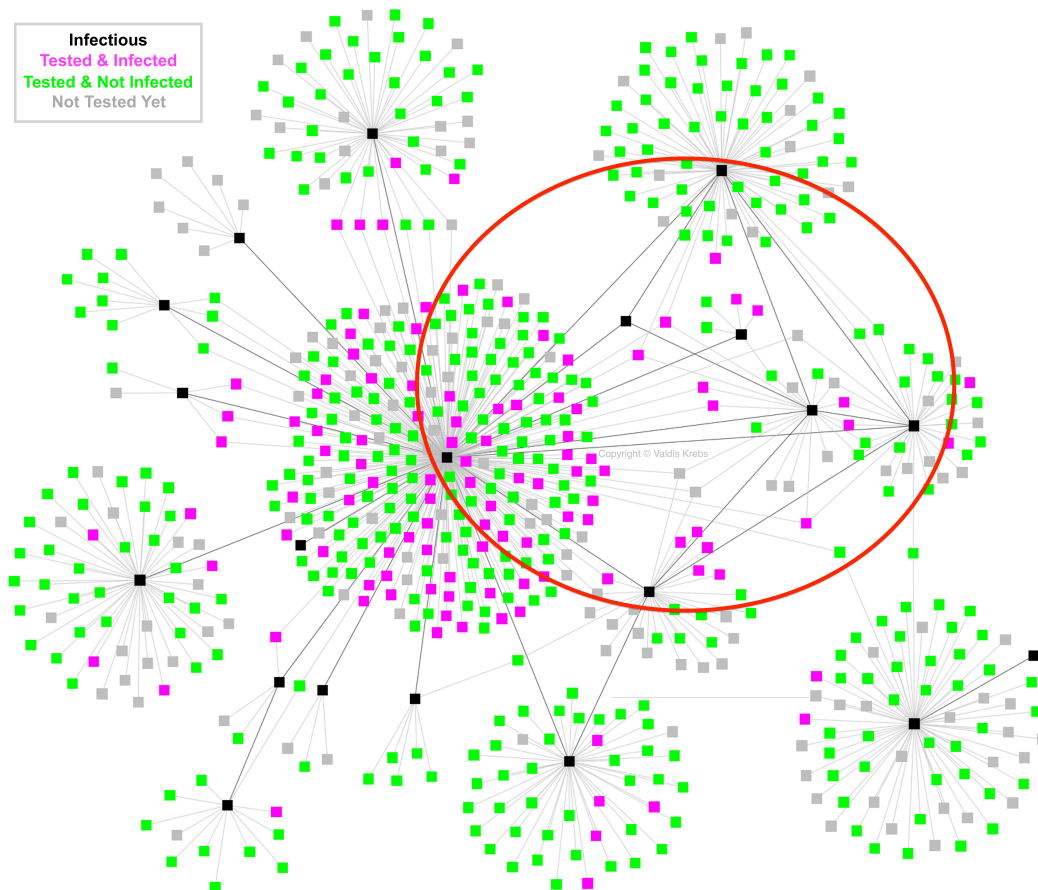
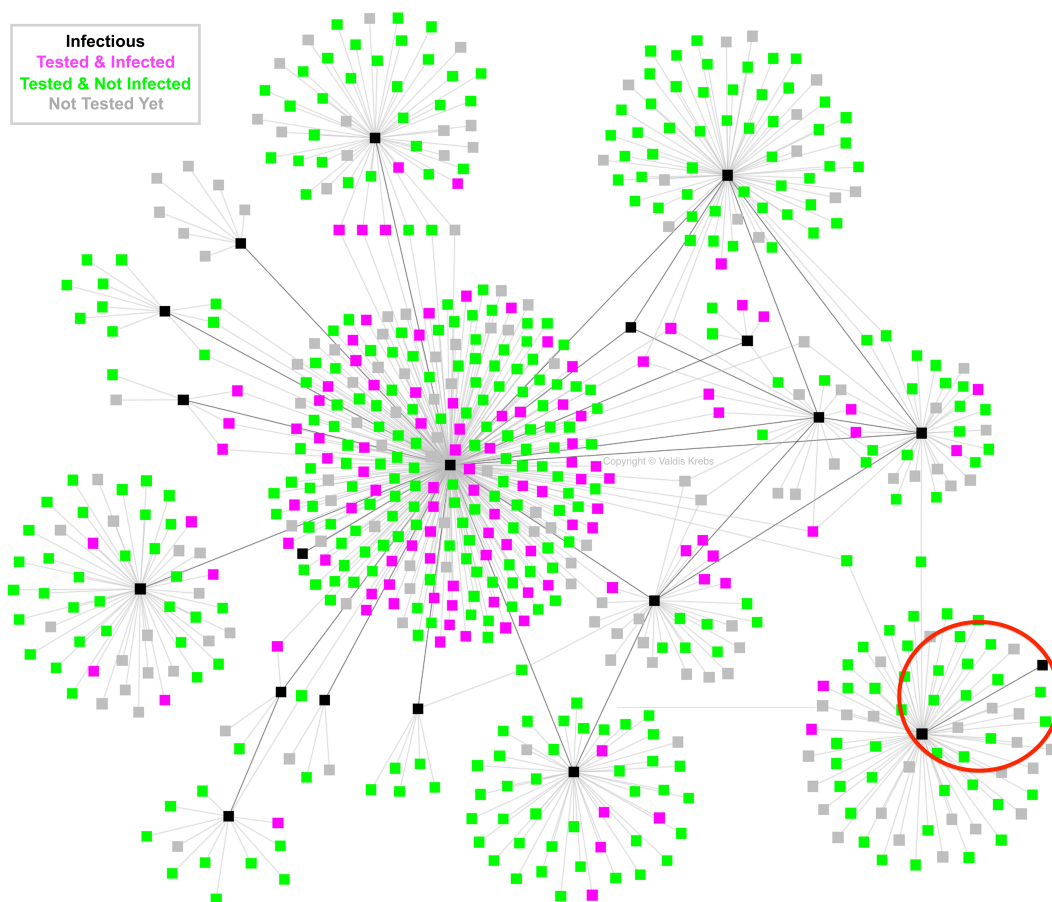


Figure 6

We examine the map in Figure 7, an interesting anomaly appears. There are two black nodes connected to each other, but not to any of the other infectious black nodes. How did these two get the disease? They do share two contacts (green nodes) with other infectious nodes but there is no line of infection to these nodes from the rest of the infectious group of black nodes. Are we missing an infectious node - a transfer point? Maybe the transfer point(s) moved out of the area? Or maybe they have died from the disease? Maybe the contact-tracing interview missed some data? There can be several reasons why the spread of the contagion is not always directly linked. These gaps in the spread of the disease need to be investigated to find possible bridges (spreaders) between communities.



**Figure 7**

Epidemiologists track a reproduction number ( $R_0$ ) of the outbreak. It is a simple measure of how a contagion spreads. It measures, *on average*, how many others does an infectious individual infect. But averages can be misleading. They reduce the variability and clumpiness of the data set to one number, which hides a lot of the interesting variability and patterns in the data set.

Looking at the network maps above we see two things immediately.

1. Some infectious people (black nodes) have more contacts than others.
2. Some infectious people spread the disease to more others (magenta/pink nodes).

The index patient definitely jumps out as being different and he skews the average greatly. But some infectious people infect 0 or 1 other person - a very low reproduction rate for the spread. An  $R_0$  of one (1) or below means that an outbreak will not happen, an  $R_0$  in the double digits means the spread will be rapid locally. Most common disease have an  $R_0$  of less than 5 - meaning the disease will spread if not controlled, but that it is manageable with a proper response - treatment, contact tracing, testing, and monitoring. It is fortunate that not everyone in this outbreak had the infection rate of the index patient. His infection rate shows what can happen if public health officials are not aware of a current disease outbreak, even if it is from just one person. Outbreaks are normally not recognized before several people in one location come down with the same disease at the same time.

Transmission Network Analysis (TNA) is best applied early in an outbreak before a mass outbreak sets in to a local area. Tracking the early cases provides great insight into how, when, and where the contagion is spreading. Data collection is difficult but always proves worthy to the understanding and management of the outbreak.

Might automated data collection aid the contact tracing process? Can we collect mobile phone *location* data of infectious persons and those in close proximity with them? This seems to have all sorts of privacy implications! But, knowing that two phones were within 3 meters for 10 minutes on X date, at Y time might help to quickly map the contact network of known infectious cases. The phone numbers could be used as both a contact ID for the database as well as a means of contacting the exposed person to notify them of possible exposure. Another problem with this automated approach might be finding too many *false positives*. After all, we have seen from the data above, and other outbreak data that close contact does not always result in infection. This is a very complex problem, but someone should be looking into the feasibility of an approach like this.

While new methods are explored, established approaches to contact tracing, aided by network analysis should be vigorously pursued in all local early outbreaks! We need to both *elongate* and *fragment* known chains of transmission, to slow their growth, and then stop the outbreak.

To summarize, we should follow this mantra:

**| TEST, TREAT, TRACK | REPEAT**

Treating, includes healing, and isolating the sick. Tracking includes, testing/re-testing, and self-isolation, of those who have had close contact with an infectious person. Stopping the spread takes a multi-pronged ongoing approach. **The outcome of this pandemic will be totally dependent on *how* we respond to it.**

*The latest version of this document can be found at:*

<http://orgnet.com/contagion.html>

## Bibliography

- [Transmission Network Analysis to Complement Routine Tuberculosis Contact Investigations](#)
- [Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases](#)
- [Review of Ferguson et al "Impact of non-pharmaceutical interventions..."](#)
- [Active Monitoring of Persons Exposed to Patients with Confirmed COVID-19 - United States, January-February 2020](#)
- [Resource estimation for contact tracing, quarantine and monitoring activities for COVID-19 cases in the EU/EEA](#)
- [Epidemiology and Transmission of COVID-19 in Shenzhen China: Analysis of 391 cases and 1,286 of their close contacts](#)
- [A Locally Transmitted Case of SARS-CoV-2 Infection in Taiwan](#)