Analyzing the Impact of Italy's COVID-19 Lockdown on National Mobility Networks

Final project for the Social Network Analysis course

Gabriele Sarti Enrico Fallacara

University of Trieste, A.A. 2019/2020

Abstract

The aim of this project report is to analyze shifts in mobility patterns during Italy's COVID-19 lockdown. More specifically, we consider the Italian mobility network for three specific days during the COVID-19 pandemic outbreak: the beginning of the **first local lockdown** in the Lodi province (25/02/2020), the beginning of **nation-scale lockdown** (03/10/2020) and the **end of nation-scale lockdown** at regional level (05/05/2020). We perform a diachronic descriptive analysis of the mobility network, apply community detection algorithms to highlight mobility shifts at a provincial level and model network connections by the mean of exponential random graphs (ERGs) to show the strong connection between mobility shifts and COVID-19 diffusion statistics. We conclude by comparing our results with the mobility reports publicly released by Google during the lockdown. Our code is available at https://github.com/gsarti/lockdown-mobility-analysis

Introduction

The nation-level lockdown implemented by the Italian Government in order to prevent the spreading of the COVID-19 disease in 2020 greatly affected the mobility patterns of roughly 60 millions of Italian citizens and residents. The Italian lockdown was composed by several subsequent phases of tightening and loosening of restrictions, with three most relevant dates arguably being:

- The **first lockdown** of ten municipalities of the province of Lodi, forming the first *zona rossa*, shortly after February 21st, 2020.
- The **national lockdown** that restricted travel at intra-municipal level, except for working necessities and family emergencies, on the evening of March 9th, 2020.
- The **reopening of inter-regional travel** that took place from May 5th, 2020.

In this report, we analyze how the Italian mobility trends have changed during the three phases defined by the aforementioned dates using a dataset containing raw movements *WAIT FOR SPECIFICATIONS* at a macro-scale level. We begin by presenting data and preprocessing steps that were taken in order to perform the analysis in Section 1. Section 2 contains a diachronic descriptive analysis of the mobility network that highlights shared and different properties of the networks through time. In Section 3, we apply highlight mobility shifts at a provincial level by the means of well-established community detection algorithms. Section 4 contains our efforts in modeling the mobility network by the means of exponential random graphs (ERGs) to show the strong connection between mobility shifts and COVID-19 spreading statistics. We conclude in Section 5 by comparing our results with the mobility reports publicly released by Google during the lockdown and summarizing our findings.

1. Data and Preprocessing

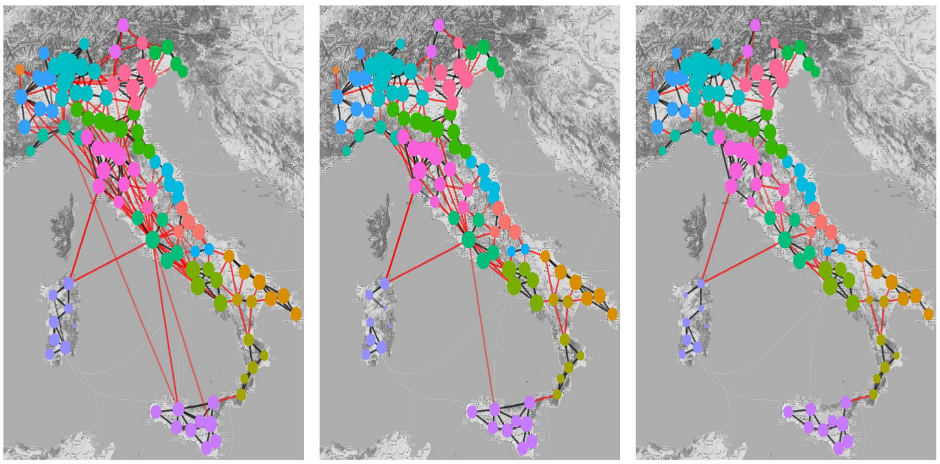
The datasets contains the estimated raw number of movements n in 3 days during the Covid crisis, filtered by country = IT. The 3 days refers to In this dataset are present different metric to estimate movements, we have selected the metric n because it represents a kind of measure of popularity.

2. Diachronic Descriptive Analysis of Mobility Networks

A first insight is given by the sum of the movements in these 3 days. We notice that mobility doesn't change much allowing intra-provincial movements, since inter-province connections have values that are much lower than intra-province ones. Thus, removing the loops, allow us to see only inter-province movements and in this case the mobility is significantly lower:

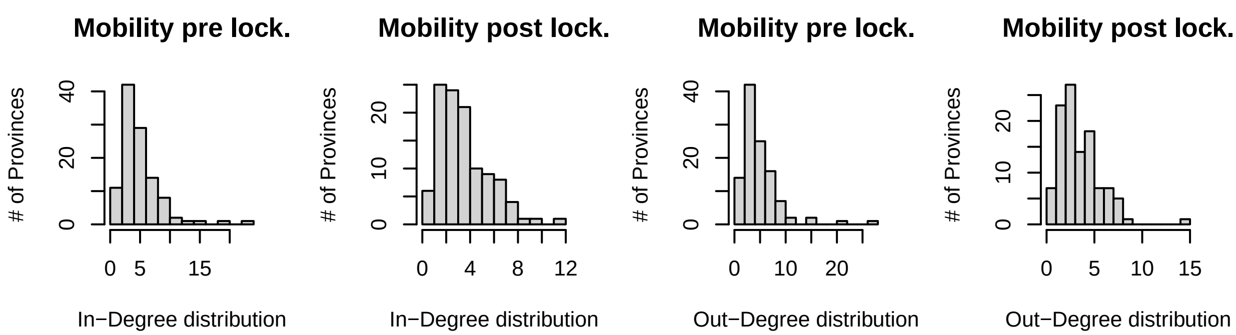
```
## [1] "Movements (inter-province): Pre 119065  -> Mid 85258  -> Post 68515"
```

A graphical representation of this phenomenon is given by the next figure.



We restrict our attention to inter-province movments for the descriptive analysis, since they capture better the variation of mobility and the results for intra-province are empirically quite similar. The graphs are **directed**, **weighted** (accoriding to the metric value) and **disconnected** (this is valid for both implementations).

From the in-degree and out-degree distribution of the 3 graphs, it is highlighted that the majority of nodes have degree between 1 and 10, with few nodes that assumes bigger values. We can notice that, for in-degree, large values tends to vanish and the distribution concentrates to small values, indicating that nodes loose their popularity through time. The same happens to the out-degree distribution, idicating that also expansiveness of nodes tends to reduce.



The **network density**, presented in Table 1, indicates, first, that the graph are **sparse**, since the very low values that assumes this quantity, and second, that the denisty reduces over time, indicating, like the in-degree and out-degree distributions, that the movements significantly decreases between the provinces through time.

In order to implement centrality measures, we have to invert the weights of our graphs. This have to be done because actual weights represent a sort of strength measure between nodes, not a distance(cost) measure.

	Pre-lock.	Mid-lock.	Post lock.
Density	0.05004170	0.03986656	0.03619683
Assortativity by degree	0.1343682	0.2511714	0.3258981
Transitivity Glo.	0.3863636	0.4052718	0.4171598
Transitivity sub. upper	0.1714286	0.2068966	0.2727273
Transitivity sub. lower	0.3148734	0.3176471	0.3448702
Transitivity k-cores	0.7959184	0.6127321	0.6113208

Table 1: Descrpitive analysis measures

After that we can calculate the **betweenness** and the **eigenvector centrality** for the graphs. Note that the closeness will be calculated later, because this measure is only meaningful in a connected graph.

The betweenness remains practically the same in all the three graphs, apart from a few small variations. Hovewer in Lazio, Umbria and Abruzzo happens an interesting thing. Before lockdown, Rome has the largest value with respect to neighboring provinces, but as the time passes, its value tends to reduce and the other neighboring provinces tend to assume higher values. This could indicate that Rome's tendency to influence movements (flows) is slightly lost in favor of its neighbors.

CITARE PLOT BTWN SUBGRAPH IN LAZIO,UMBRIA E ABRUZZO

Eigenvector centrality shows that there is a loss of influence over the whole network by provinces in the provice of northern and southern Italy, while in central Italy the situation remains almost completely unchanged, despite a slight decrease in values.

CITARE PLOT EIGE_VECT SUBGRAPH IN TUTTA ITALIA

For what it concerns about **assortativity by degree**, we calculate this measure for the graphs and, as reported in Table 1, we can see that in the mobility pre-lockdown we have a small positive value of 0.134 that indicates the sligth tendency of nodes of similar degree to connect with each others. An interesting fact is that this measure increases through time, reaching a value of 0.326. This could indicate that the movements have concentrated more between provinces of the same degree, reducing the movement from "small" to "large" ones.

The **global clustering coefficient**, that refers to netowrk cohesion analysis, is calculated:

- **Globally**
- on **subgraphs** in which all the nodes have a weight that is above or below a fixed **threshold**.
- in **k-cores components** of original graphs.

As reported in Table 1, the **transitivity** for the whole network tends to remain the same, except for a small increase that isn't significant. Thus, to inspect better this property, we calculate this measure on two subgraphs obtained placing a $threshold_w = 500$ on weights. The value that we choose is based on the distribution of weights in the network, noticing that most part of nodes have weights that go from 0 to 500. As we can notice from the results presented in the table, for the subgraphs of nodes below the threshold, the transitivity, that is smaller with respect the global value, remains again pratically the same, instead for the other subgraph there is a significant increment of 0.10. This could mean that big provinces tends to increase their connections through the three different phases of lockdown. We have also calculated this index on k-cores of the graph, selecting as k the minimum value between the three coreness value, resulting in $k = 8$ (notice that only the graph referred to the mobility pre-lockdown has a value of k bigger than the selected, with the coreness equal to 10). The transitivity calculated on these maximal subgraphs has a significant decrease of 0.18, as we expected, indicating that these subgraphs have a loss in terms of cohesion through the time.

After this first descpritive analysis we can deduce that the decrease of movements caused by lock-down measures causes a loss of influence of the nodes in the whole network, with an increasing tendecy of similar degree nodes to connect with each others, indicating that movements from small provinces to large ones decrease, in favor of smaller and closer provinces, in terms of distance. The opposite happens for big provinces, movements from bigger to smaller reduce in favor of movements between big ones.

3. Analyzing Mobility Communities

We limit ourselves to the analysis of `inbound_strength`, i.e. the sum of the weights of inbound edges to each node. Intuitively, it can be regarded as a proxy for the importance of the node inside the network. Figure N shows that movements across provinces follow a power law that is typical for those types of networks, with very few provinces accounting for most of the movements. We see from the histograms that the value of `inbound_strength` is much higher when accounting for intra-provincial movements. This is also evident in the visualizations of Appendix X, where this value is used to weight edge widths. While total movements don't seem to be affected, we can see a drop of roughly 30% in inter-provincial movements between February 25th and March 10th, with a further 13% decrease in mobility between March 10th and May 5th. These percentages match our intuition that inter-provincial mobility suffered the most from inter-regional lockdown measures.

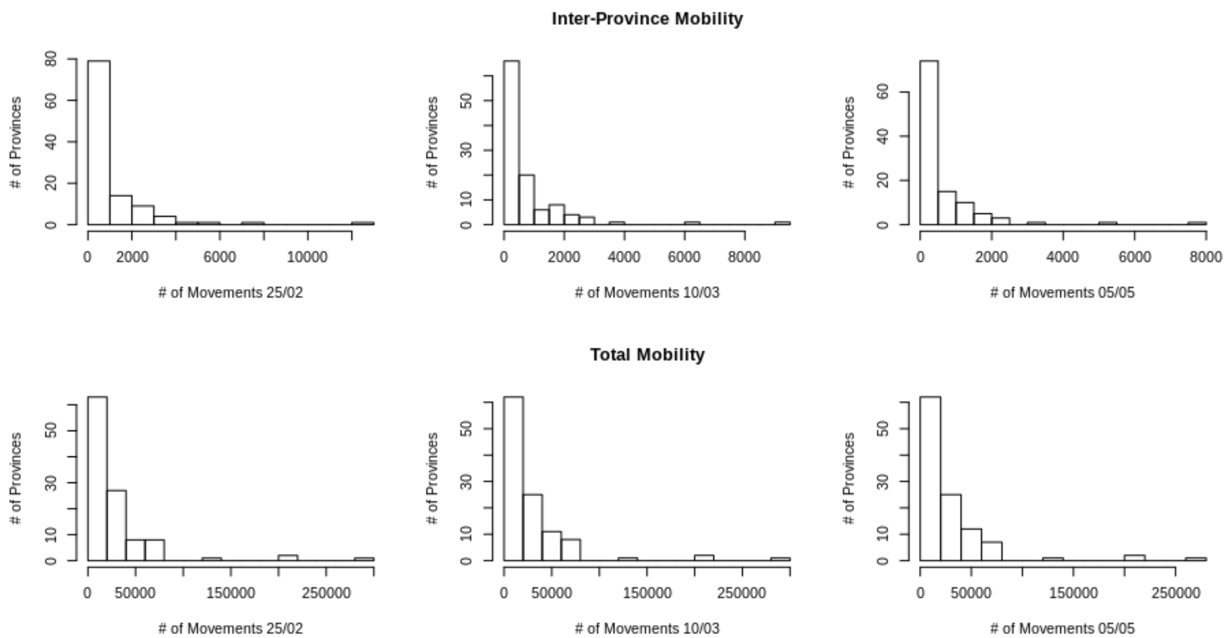


Figure 1: Mobility histograms representing the distribution of total inbound movements across provinces at various stages of the Italian lockdown.

We proceeded to compute and visualize the coreness of the network, as shown in Appendix X. As expected, removing intra-provincial movements (looping edges) doesn't affect at all the coreness structure. From the coreness plots structure it is evident how north-eastern, southern and insular provinces have a much lower number of connections with respect to central (e.g. Tuscany) and north-western (e.g. Lombardy, Piedmont) ones. Calabrian and Sardinian provinces have the lowest number of connections.

In order to account for the intrinsically weighted structure of our networks, we developed a **weighted version of the coreness algorithm** that groups vertices in bins of fixed length based on their weights and uses bin membership as a proxy for weighted coreness. In our specific case, we used a bin size of 500 raw movements for the network without intra-provincial movements, and 10'000 for the network containing all movements. The details of our implementation are presented in Algorithm 1. We see from the results shown in Appendix X that our approach effectively highlights the centrality of northwestern provinces when we account only for inter-provincial movements, while it shows a strong correlation with population counts when accounting for all external and internal movements in the province.

Algorithm 1: Weighted Coreness

Input: Network $\mathcal{G}(V_G, E_G)$, strenght bin size b
Set current weighted k-core subgraph $\mathcal{K}(V_k, E_k) \leftarrow \mathcal{G}$
Set max inbound strength $S \leftarrow \max(\text{strength}(V))$
Set index $i \leftarrow 1$
for *min in* $b, 2b, 3b, \ldots S$ **do**
    Create bin graph $\mathcal{B}(V_B, E_B) \leftarrow$ Induced subgraph($\mathcal{K}$, strength$(V_k) \leq min$)
    if $|V_B| > 0$ **then**
        Set *wcore* of V_G that were not filtered in $\mathcal{B}$ to i
        $\mathcal{G}(V_B \cap V_G)[wcore] \leftarrow i$
        $i \leftarrow i + 1$
    end
    Set current weighted k-core subgraph to the complement of bin graph
    $\mathcal{K} \leftarrow$ Induced subgraph($\mathcal{K}$, strength$(V_k) > min$)
end
Output: Network $\mathcal{G}(V_G, E_G)$ with *wcore* vertex attributes

---

For the second part of the community analysis, we proceeded only with the network of inter-provincial movements since the presence of looping edges wasn't relevant for detecting communities inside the network. We inspected the component structure of the network in the three analyzed dates, finding that a single giant component comprising almost all the provinces, with few isolated nodes left aside in Southern Italy, was present at all times. This isn't surprising, as we expected some level of communication at a national level at all times for business and family-related reasons despite the lockdown.

In Appendix X we show the similarity between the structure of cliques through time and the components found by considering network actors with weighted coreness rank > 2, i.e. with at least 1000 inter-provincial outbound movements at all times. This highlights the fact that most tight-knit communities are also the ones having the higher number of connections in general, reflecting the principle of preferential attachment which is commonly verified in many real-world social networks. This fact is also verified by the presence of a **small-worldness index** above 3 at all times, and a global transitivity that is significantly above chance. Both cliques and weighted coreness components highlight the centrality of Milan and its neighbouring provinces, Tuscany provinces and the Rome-Naples macro-region. Weighted coreness components further provide additional information by highlighting the high connectivity of Veneto and representing well the drop in connectivity during the lockdown.

To conclude the section, we performed some community detection experiments leveraging two well-established algorithms. The **Girvan-Newman algorithm** finds communities by progressively removing edges based on their betweenness. Since edge weigths are interpreted as distances by the algorithm, we converted our raw movement weights w following an approach suggested in the literature (Özyer, Bakshi, and Alhajj 2019, 66):

$$w' = \frac{1}{w} \times 10'000$$

The **Label propagation algorithm**, instead, iteratively sets node community memberships based on the weighted most common membership across neighbouring nodes until convergence. This approach doesn't require inverting raw movements weights w since larger edge weights correspond to stronger connections. Results produced by the two methods are presented descriptively in Table X and visually in Appendix X. We could not apply standard multi-level modularity optimization methods given the directed nature of our network.

The label propagation algorithm appears to be more effective at modeling the community structure of the networks at various times. Communities found with label propagation closely approximate regional

	Girvan-Newman			Label Propagation		
	25/02	**10/03**	**05/05**	**25/02**	**10/03**	**05/05**
# communities	36	60	7	22	22	24
modularity Q	.39	.25	.39	.51	.57	.59
δ_{int}	.19 (.38)	.12 (.31)	.54 (.37)	.78 (.21)	.68 (.25)	.73 (.23)
δ_{ext}	.07 (.05)	.08 (.04)	.01 (.01)	.04 (.03)	.03 (.01)	.03 (.01)
Gini index	.96 (.18)	.97 (.14)	.66 (.37)	.78 (.22)	.76 (.22)	.80 (.21)

Table 2: Values of the metrics computed using two community detection approaches. Values for intra-cluster density δ_{int}, inter-cluster density δ_{ext} and Gini index are presented in the *mean (stdev)* format.

membership at all times, a fact that matches the empirical presence of regional hubs of communication across the Italian territory. It is interesting to note that both algorithms correctly highlight a decrease in inter-community edges over time, which is also made evident by the drastic reduction in red inter-cluster edges in the plots of Appendix X.

4. Modeling Mobility Patterns

In order to implement ERGMs on our networks, we decide to insert some **auxiliary data** that could allow us a better model fitting. These data came from two different dataset:

- NOME DATASET, that contains informations about **population**, **population density**, **surface** and **municipalities**, aggregated at province level.
- COVID-19 italian dataset, that contains informations about the numer of **covid cases** at province level and number of **covid deaths** at regional level.

These variables are chosen because they can be significant in our analysis, as they could influence the movements between the different regions and/or provinces.

After adding these variables, we procede with the specification and the estimation of three different models, one for each graph. The general idea in order to specify the model is inspired by the stepAIC() function that performs backward model selection by starting a "maximal" model, which is then trimmed down: starting from the maximal model, we remove iteratively the less significant variable at each step, until we obtain a model in which all the variables are significant, then we select the one that has the lowest AIC value.

The three final models are:

```
model_pre <- ergm(gc_pre ~ edges + mutual + nodematch('region', diff=F) + nodecov('ideg')
                           nodecov('out_strength'))
model_mid <- ergm(gc_mid ~ edges + mutual  + nodematch('region', diff=F) + nodecov('odeg'))
model_post<-ergm(gc_post ~ edges + mutual+ nodecov('covid_deaths') + nodecov('ideg')
                           nodematch('region', diff=F) +nodecov('eig_vect_centr'))
```

All the three models use informations about the number of edges, mutuality (number of pairs of actors i and j for which $(i \to j)$ and $(j \to i)$ both exist) and homofily, given by the match by region, of the networks. They differ instead in the covariates they use: the first two models don't use statistics relative to COVID-19 and this make sense, because on those dates the data relating to the epidemic did not yet show its real extent on the territory (low values of covid cases and deaths). Instead, the third model uses the covid deaths, using also the eignevector centrality that reflects the node's centralization property.

We notice an important thing: these 3 models shows a sort of *performance decay*, in the sense that the first model, according to Goodness-of-fit diagnostics performs pretty well, while the second and third show worse and worse performances, even if for all the three models the MCMC diagnostics are good and they include all the relavant predictors. This could be a hint that three single ERGM models are not suitable for our purpose, thus considering a temporal network data analysis seems more appropriate for our problem, in which we have a graph that changes through three different moments of time.

CITARE PLOT GOF AND/OR MCMC FOR THE THREE BASIC MODELS

5. Discussion and Conclusion

Appendix

A. Coreness Plots for Feb. 25th, 2020 Mobility Network

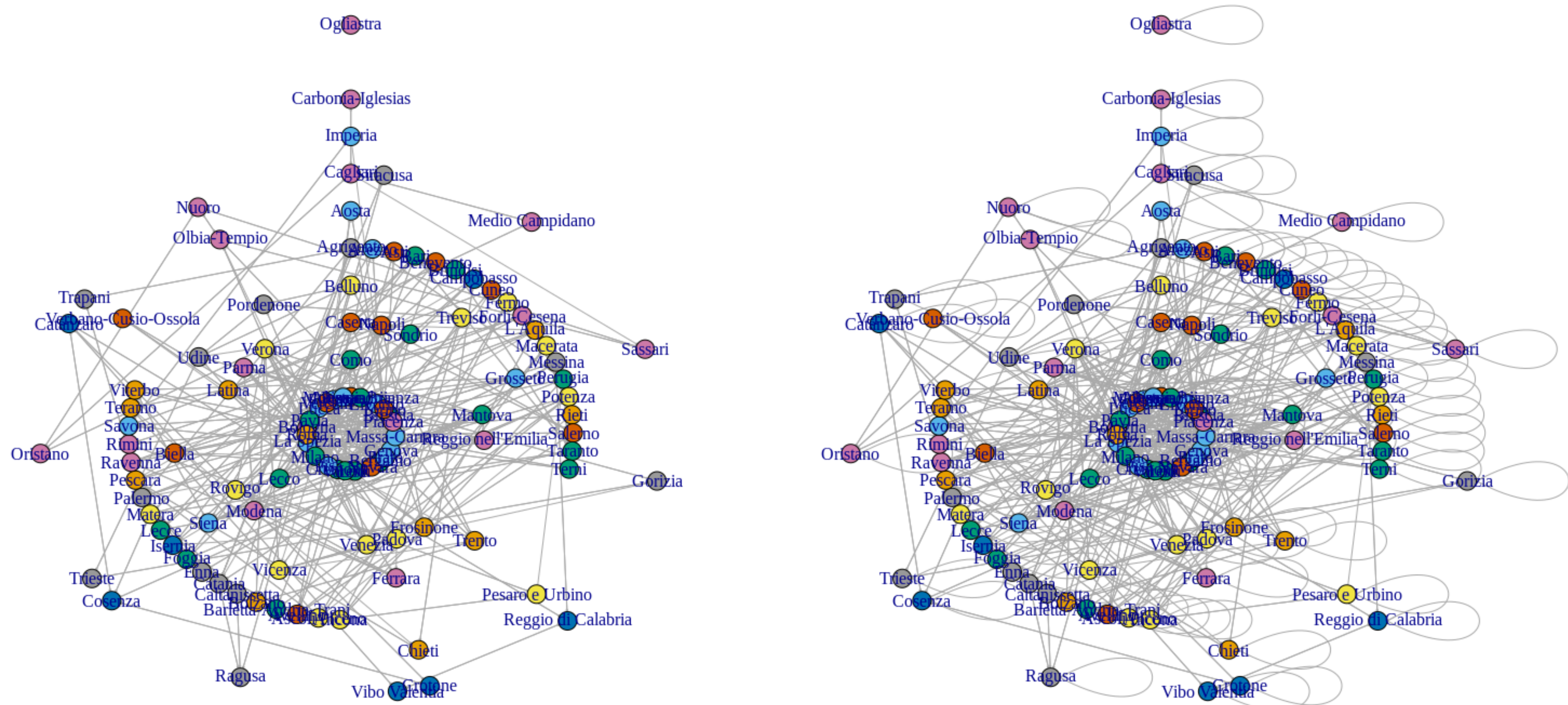


Figure 2: Finding the .rproj file for Chapter 3

B. Weighted Coreness Plots for Feb. 25th, 2020 Mobility Network

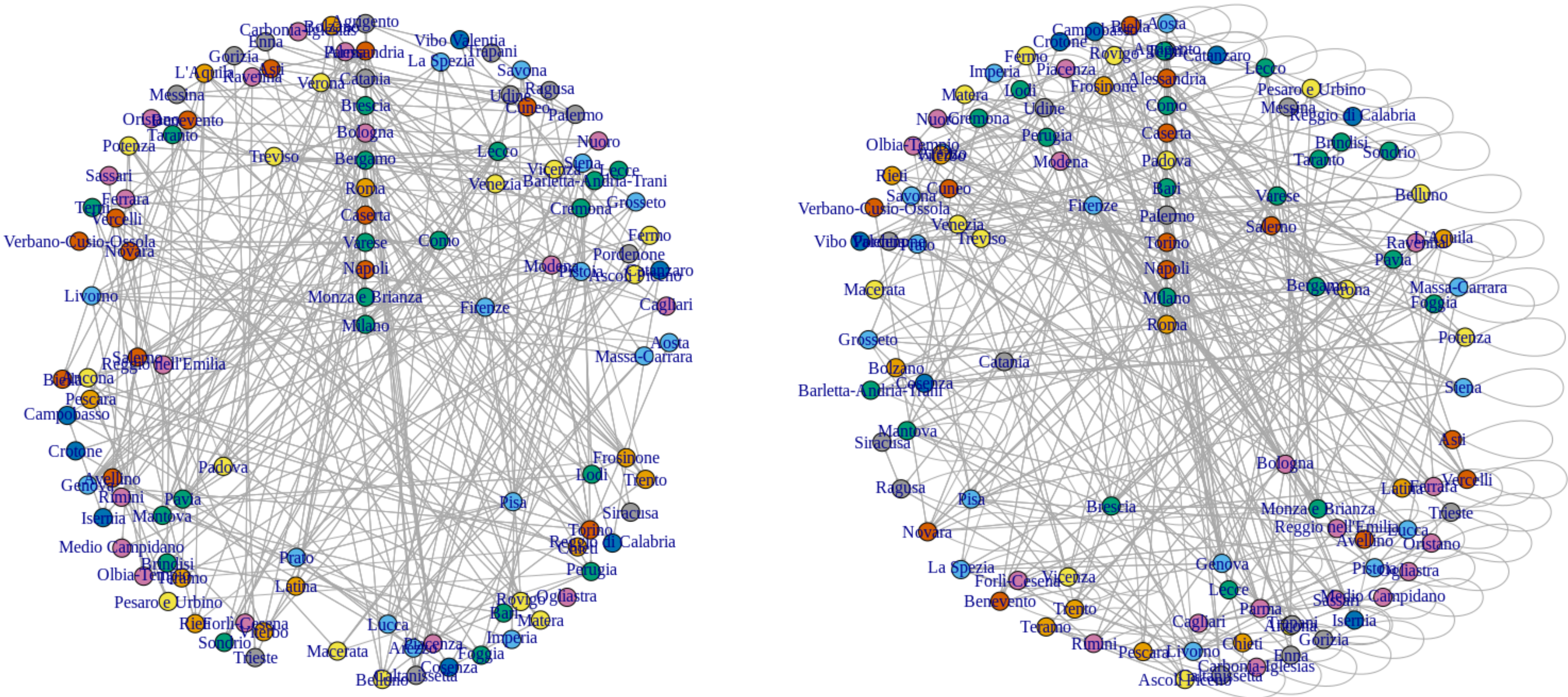


Figure 3: Finding the .rproj file for Chapter 3

C. Connected Components and Cliques for the Three Lockdown Phases

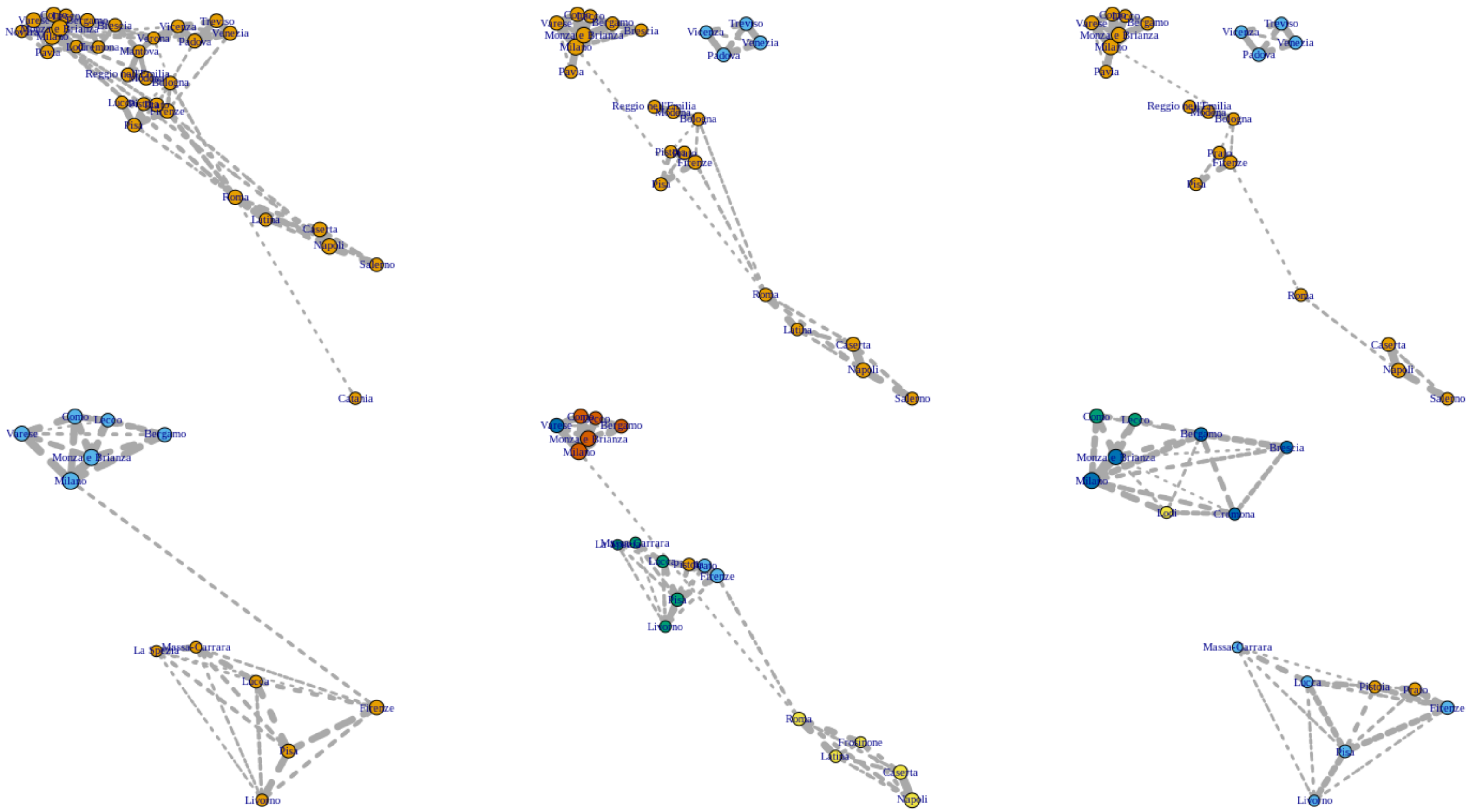


Figure 4: Finding the .rproj file for Chapter 3

C. Community Detection using Girvan-Newman Algorithm and Label Propagation

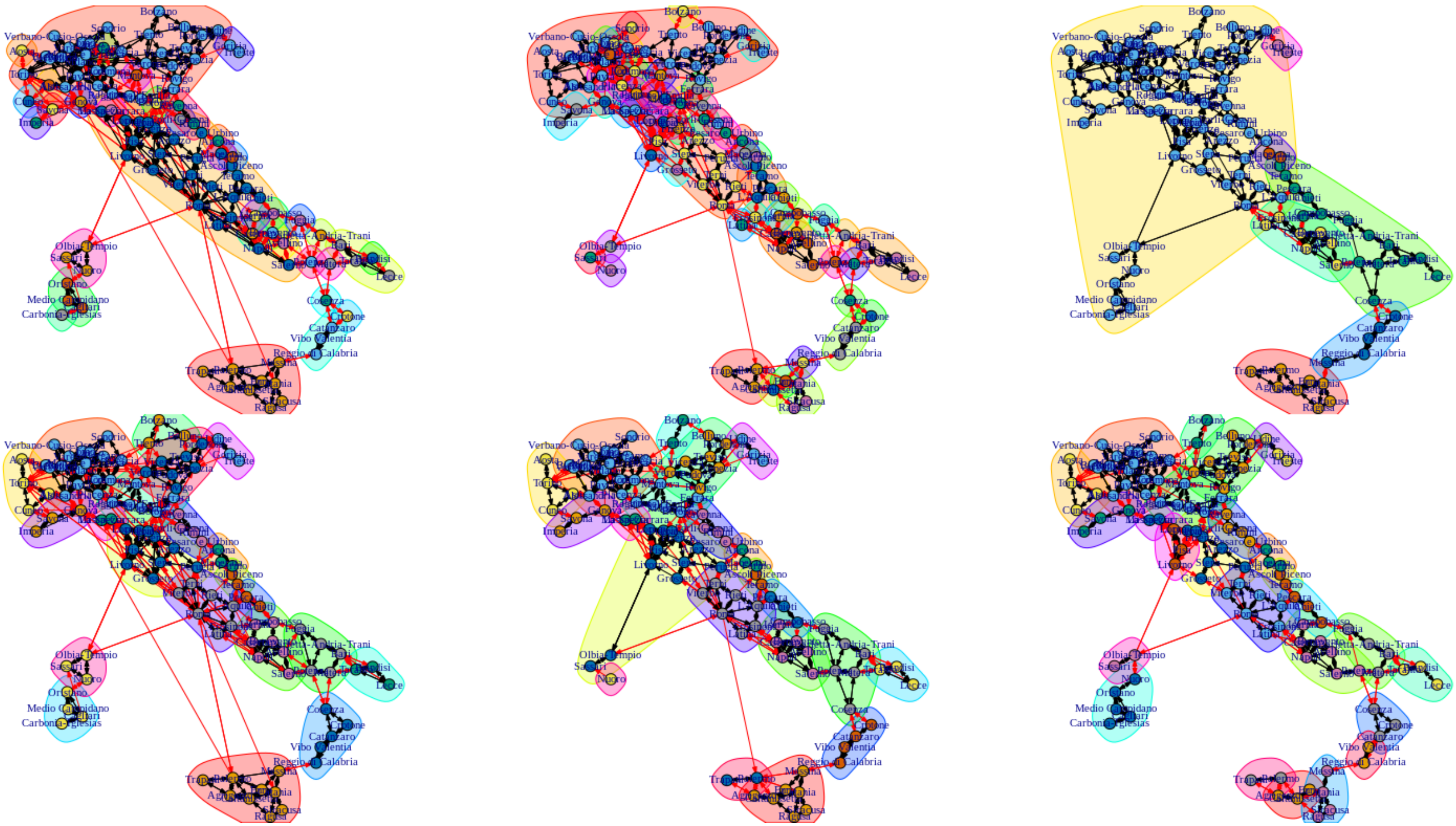


Figure 5: Finding the .rproj file for Chapter 3

D. Regions of middle Italy where the size of each vertex is proportional to betweenness

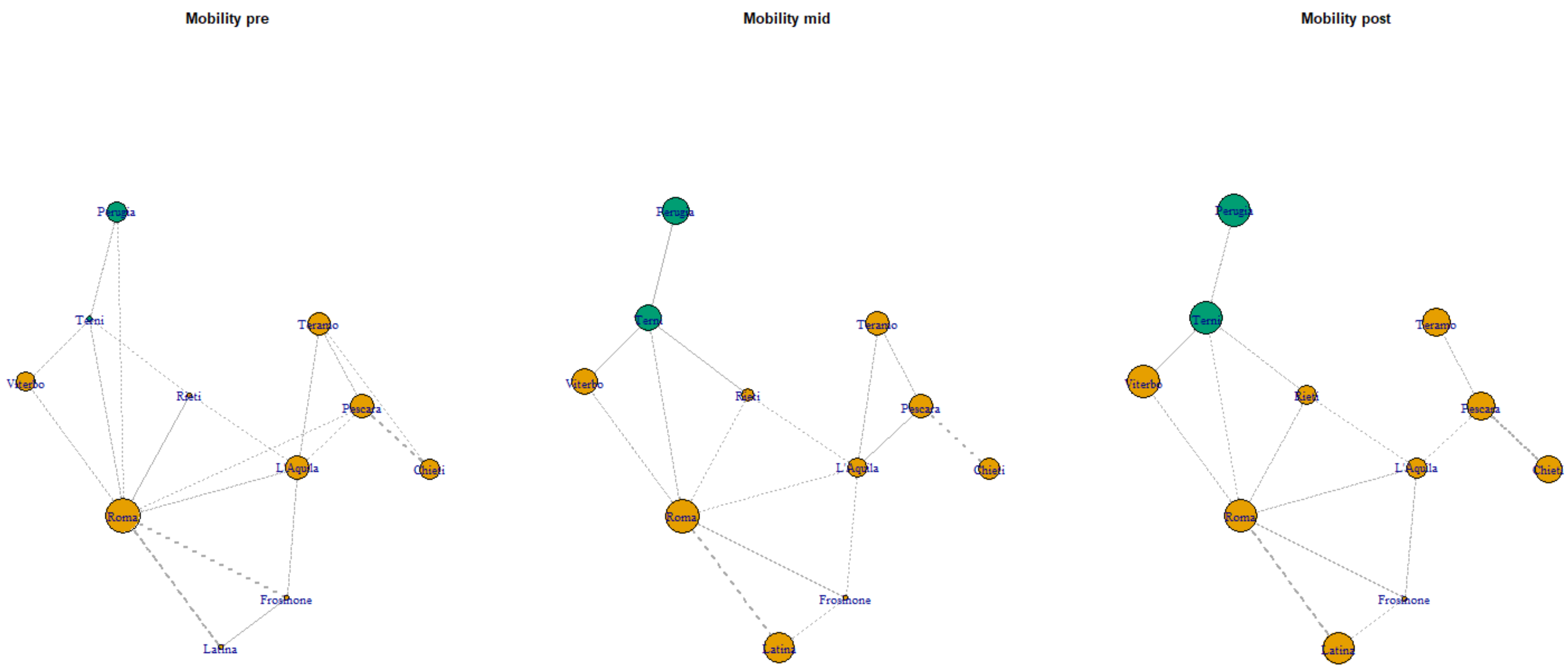


Figure 6: Finding the .rproj file for Chapter 3

E. Network graphs where the size of each vertex is proportional to eigenvector centrality

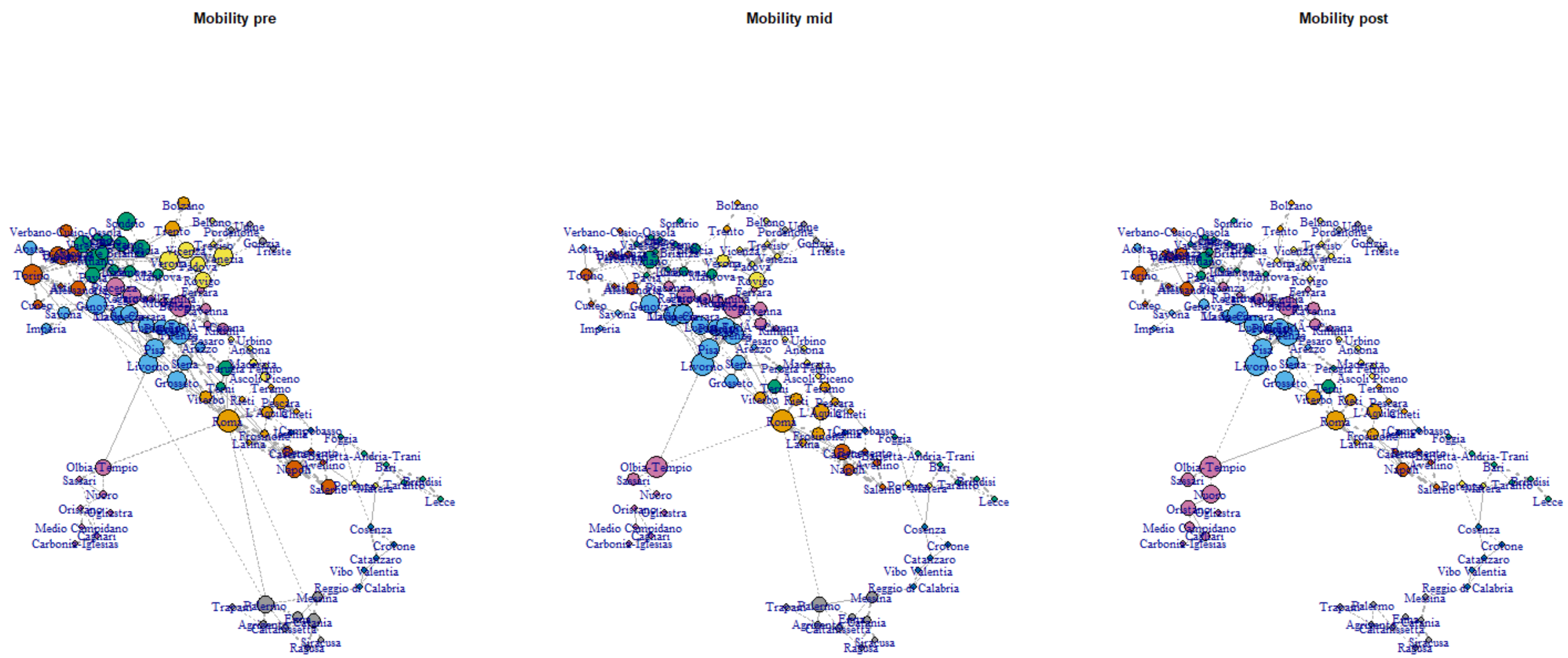


Figure 7: Finding the .rproj file for Chapter 3

References

Özyer, Tansel, Sambit Bakshi, and Reda Alhajj. 2019. *Social Networks and Surveillance for Society.* Springer.