Modelling interactions among offenders: A latent space approach for interdependent ego-networks

Isabella Gollini
University College Dublin, Ireland

Alberto Caimo
Technological University Dublin, Ireland

Paolo Campana
University of Cambridge, UK

Abstract

Illegal markets are notoriously difficult to study. Police data offer an increasingly exploited source of evidence. However, their secondary nature poses challenges for researchers. A key issue is that researchers often have to deal with two sets of actors: targeted and non-targeted. This work develops a latent space model for interdependent ego-networks purposely created to deal with the targeted nature of police evidence. By treating targeted offenders as egos and their contacts as alters, the model (a) leverages on the full information available and (b) mirrors the specificity of the data collection strategy. The paper then applies this approach to analyse a real-world example of illegal markets, namely the smuggling of migrants. To this end, we utilise a novel dataset of 21,555 phone conversations wiretapped by the police to study interactions among offenders.

1 Introduction

Every day a considerable number of interactions take place outside the realm of legal frameworks. Individuals across the world produce and trade a variety of illegal products and services, ranging from drugs to counterfeit goods, stolen products or illegal border crossings (Campana and Varese, 2018). This multitude of interactions constitutes the backbone of illegal markets. Studying such interactions is a crucial task if we are to understand how illegal activities are organised and how illegal actors operate. Morselli (2013), Faust and Tita (2019) and Campana and Varese (2020) offer a review of the questions scholars have been interested in: these range from modelling exposure to violence to understanding co-offending patterns in illicit networks of various kind as well as exploring the dynamics internal to illegal organisations such as organised crime groups and gangs. Yet, studying illegal interactions can be a rather challenging endeavour.

Scholars have pointed to a number of issues that researchers face when studying hard-to-reach and hidden populations (Atkinson and Flint, 2001). Faust and Tita (2019), Diviák (2019) and Campana and Varese (2020) offer a comprehensive review of the challenges and pitfalls that researchers might encounter when using social network analysis in criminological research, including illegal markets and organised crime. One major issue is the availability of data – particularly in the context of quantitative network research. To overcome this problem, scholars are increasingly relying on law enforcement data (see, among others, Natarajan (2000); McGloin (2007); Morselli (2009); Papachristos (2009); McGloin and Piquero (2010); Malm and Bichler (2011); Campana (2011, 2016); Grund and Densley (2012); Bright et al. (2012); Schaefer (2012); Papachristos et al. (2012); Varese (2013); Campana and Varese (2013); Papachristos et al. (2015); Calderoni et al. (2017); Bright et al. (2018)).

However, such data come with limitations. For example, they might be influenced by the level of enforcement, policing priorities, recording practices as well as resource constraints (Morselli, 2009; Malm and Bichler, 2011; Campana and Varese, 2012; Calderoni, 2014; Faust and Tita, 2019; Campana and Varese, 2020). We refer to Faust and Tita (2019) and Campana and Varese (2020) for a broader discussion of such limitations. In this paper, we focus on the secondary nature of such data. Normally, researchers have no input in designing the data collection strategy adopted by law enforcement agencies and thus need to work within the boundaries set by the agency. The secondary nature of the evidence makes it difficult to control for, alas, errors and missing data (Malm et al., 2008). A further implication is that researchers are constrained by the sampling strategy adopted by the law enforcement agency in the first place. This is a key issue that has far-reaching modelling implications.

For example, during an investigation, police normally target a sub-set of individuals and then collect information about additional individuals connected to the targeted ones. This creates two sets of actors: the targeted individuals and the non-targeted individuals. Police investigations can be seen as a specific type of link-tracing design in which referrals are unwittingly provided by the 'respondents' (Heckathorn and Cameron, 2017). During an investigation, new targeted individuals may be added, but those will inevitably bring in a new set of non-targeted individuals. It is almost inevitable, due to the nature of police investigations, that we end up with two sets of actors: those who have been directly targeted and those who have entered the dataset by virtue of being connected to the targeted one. In theory, one could envisage a situation in which (a) all the previously non-targeted individuals are targeted before the end of the investigation and (b) all their contacts are already included in the dataset: this situation, however, never occurs in reality. (Incidentally, the fact that investigators make decisions on whom to target is a different issue, which is treated in Campana and Varese (2020)). The targeted nature of police investigations is universal across jurisdictions.

Researchers who wish to work with police records face a similar problem if the data were extracted using a targeted extraction strategy (Campana and Varese, 2020). This strategy normally consists in selecting a set of actors based on certain characteristics, e.g. being part of an organised crime group or a gang, and then extract all the alters connected to the initial set of actors. If the 'alter-alter' relations are not *directly* extracted, then we are in a situation in which part of the actors are directly targeted and part are not (this was the case, for example, of the dataset used in Ouellet et al. (2019) or in Campana and Varese (2020)). Whether it is due to investigative practices or to the type of extraction strategy adopted, the targeted nature of police evidence has an impact on the structure of the data made available to researchers. The latter are then confronted with a difficult issue related to the treatment of such data as the likelihood of appearing in the network is not the same for the targeted and the non-targeted individuals (see Campana and Varese (2013), Bright et al. (2018), Diviák (2019), Campana and Varese (2020); this is similar to the 'spotlight effect' discussed by Smith and Papachristos (2016)). If researchers focus on the targeted individuals only, they will disregard a very large amount of potentially valuable information. How can we then consider the evidence on both targeted and non-targeted individuals in a way that takes into account the specificity of the data collection strategy *and* minimises the amount of information that we disregard?

Varese (2013), Smith and Papachristos (2016) and Campana (2018) have adopted an indirect strategy by sub-setting the initial dataset and then running robustness checks. Combining years of experience in our respective fields, in this paper we offer a novel solution that *directly* models the specificity of targeted evidence. This modelling strategy is based on a latent space framework (Hoff et al., 2002; Handcock et al., 2007; Gollini and Murphy, 2015; Rastelli et al., 2016) for interdependent ego-networks. We suggest treating all targeted individuals as egos and all non-targeted individuals as alters. Our approach consists in assuming that the latent positions of the egos will be jointly determined by the ego-ego and ego-alter connectivity structure so that the closer the positions of two egos in the latent space the higher the probability that the two egos have a link between them and share common alters.

In criminology, latent space models have been applied only recently to study the heroin drug flows among countries (Berlusconi et al., 2017). We advance this line of work by offering the first application of latent space models to actor-level patterns of interactions; further, we present a tailored model to capture interdependent egos. We suggest that this approach can be fruitfully applied to answer a number of research questions related to illegal markets and offenders' behaviour. While the formulation of such questions necessarily depends on the specific content of the evidence that a researcher can rely upon, potential examples relate to the study of the structure of illegal markets and the interactions underpinning those. A researcher can answer questions related to the identification of clusters of dense interactions – what we could call "criminal proximity" – as well as the opposite notion of "criminal distance" (in a latent space approach, relative distances between actors are meaningful). One can compare distance-based clusters with, for instance, attribute-based clusters (e.g. organised crime or gang membership). Further, one can identify actors with a high degree of equivalence from a criminal market perspective, thus flagging a potential for a quick replacement if one of the pair is arrested. Crucially, close association (clustering) and criminal distances are calculated taking into account not only direct interactions among (targeted) offenders but also their alters' profile. If the evidence allows, this approach can be applied to *jointly* explore the supply-side and the demand-side of an illegal market, for instance in situations in which sellers have been targeted and customers can be identified among the non-targeted population. Drug dealing comes to mind: for instance, we can model association and criminal distance between dealers *also* taking into consideration their customers' profile.

In this paper, we present an application of our approach to study a topical issue in contemporary societies: the smuggling of migrants. We will do so by relying on a novel data set of real-world wiretapped phone conversations among human smugglers that we obtained from the Italian police.

The paper proceeds as follows: the next Section discusses the latent space model for interdependent ego-networks. Section 3 introduces the data for this study and their structure. Section 4 presents the results of the models and Section 5 offers a further analysis of the estimated link probabilities. Section 6 concludes.

2 Interdependent ego-networks

The evidence collected during police investigations usually generates a data structure akin to a collection of interdependent ego-networks. Some individuals are normally placed under surveillance, for instance they have their phone lines wiretapped. Other individuals, on the other hand, are included in the evidence by virtue of having been connected to someone under direct surveillance, for instance through a phone call they have made or received. A targeted extraction from police records generates a similar data structure. We interpret the first set of actors–the targeted ones–as egos. The second set of actors–the non-targeted ones–are the alters.

More formally, let N be the number of observed egos and $\mathbf{Y}$ be the $N \times N$ adjacency matrix containing the relational information between them, with entries $y_{ij} = 1$ if there is a tie between ego i and ego j and $y_{ij} = 0$ otherwise. Let M be the number of observed alters and $\mathbf{X}$ be the $N \times M$ incidence matrix encoding presence or absence of an edge between egos and alters, with entries $x_{ik} = 1$ if there is a tie between ego i and alter k and $x_{ik} = 0$ otherwise. Egos can be connected to the same alter l if $x_{il} = x_{jl} = 1$. See Figure 1 for a graphical representation of the relational structure of interdependent ego-networks.

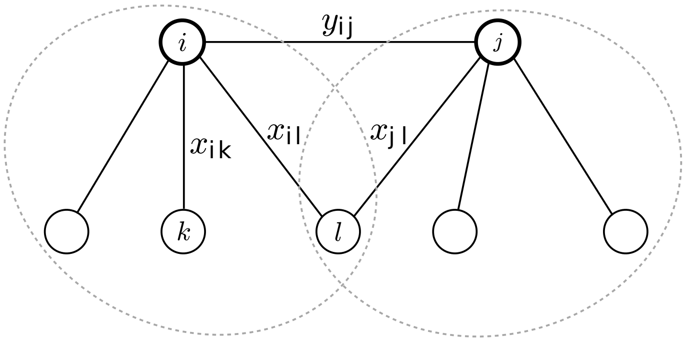

Figure 1: Example of two interdependent ego networks: two egos i and j can be connected through y_{ij} and they can also share a common alter (l).

2.1 Latent space model

The latent space modelling approach described provides an interpretable model-based visual representation of the network connectivity structure as it takes into account several relational properties. The latent space model proposed by Hoff et al. (2002) assumes the existence of a D-dimensional latent space where nodes are positioned according to their probability of being connected. The nodes' positions are determined by a logistic regression model that assumes that the shorter the latent distance between two nodes, the higher the probability that those two nodes are connected.

Several distance metrics have been proposed in the literature according to the various types of link relations (see, for example, Hoff (2005, 2009)). Gollini and Murphy (2015) proposed to use the squared Euclidean distance for undirected networks instead of the commonly used Euclidean distance (Hoff et al., 2002) for two main reasons: firstly, it allows one to visualise more clearly the presence of nodal clusters by giving a higher probability of a link between two close nodes in the latent space and lower probabilities to two nodes lying far away from each other. Secondly, it makes the model need fewer approximation steps for the variational estimation procedure which provides a very fast estimation method for large networks (see Section 2.3 for more details). This is particularly helpful when dealing with large-scale police evidence.

The relational structure of the ego-ego network $\mathbf{Y}$ can be captured by a latent space model with squared Euclidean distance:

$$p(\mathbf{Y} \mid \mathbf{Z}, \alpha) = \prod_{i \neq j}^{N} p(y_{ij} | \mathbf{z}_i, \mathbf{z}_j, \alpha) = \prod_{i \neq j}^{N} \frac{\exp(\alpha - |\mathbf{z}_i - \mathbf{z}_j|^2)^{y_{ij}}}{1 + \exp(\alpha - |\mathbf{z}_i - \mathbf{z}_j|^2)}, \tag{1}$$

where the density parameter and the latent positions are respectively $\alpha \sim \mathcal{N}(\xi_\alpha, \psi_\alpha^2)$, and $\mathbf{z}_i \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_D)$ and $\xi_\alpha, \psi_\alpha^2, \sigma^2$ are fixed parameters.

2.2 Latent space model for interdependent ego-networks

The main aim of the latent space model for interdependent ego networks is to visualise the position of egos (and alters) based on both the ego-alter and ego-ego connectivity structure in a unique interpretable way. To do so we assume the existence of a D-dimensional latent space on which both egos and alters lie. To take into account the dependence structure within the ego networks we first define the probability that an ego i and an alter k are connected as

$$p(x_{ik} \mid \mathbf{z}_i, \mathbf{w}_k, \beta) = \frac{\exp(\beta - |\mathbf{z}_i - \mathbf{w}_k|^2)^{x_{ik}}}{1 + \exp(\beta - |\mathbf{z}_i - \mathbf{w}_k|^2)},$$

where $\mathbf{z}_i$ and $\mathbf{w}_k$ are the latent positions of ego i and alter k respectively. Assuming conditional dyadic independence given the latent positions we have that the overall probability of observing the incidence matrix $\mathbf{X}$ can be written as

$$p(\mathbf{X} \mid \mathbf{Z}, \mathbf{W}, \beta) = \prod_{i=1}^{N} \prod_{k=1}^{M} p(x_{ik} \mid \mathbf{z}_i, \mathbf{w}_k, \beta),$$

where $\beta \sim \mathcal{N}(\xi_\beta, \psi_\beta^2)$ is a baseline density parameter, $(\mathbf{Z}, \mathbf{W}) \overset{iid}{\sim} \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_D)$ represent respectively the latent positions of egos and alters in the latent space, and $\xi_\beta, \psi_\beta^2, \sigma^2$ are fixed parameters.

The dependence structure between the ego networks is then captured by the latent space model defined by $p(\mathbf{Y} \mid \mathbf{Z}, \alpha)$ in Equation 1. Therefore the likelihood of the latent space model for interdependent ego-networks can be written as

$$p(\mathbf{Y}, \mathbf{X} \mid \mathbf{Z}, \mathbf{W}, \beta, \alpha) = p(\mathbf{Y} \mid \mathbf{Z}, \alpha)\, p(\mathbf{X} \mid \mathbf{Z}, \mathbf{W}, \beta) \tag{2}$$

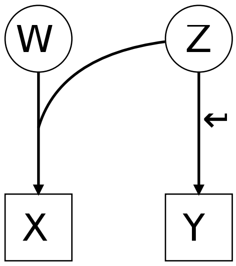

Figure 2: Graphical representation of the latent space model for interdependent ego-networks.

and its graphical representation is displayed in Figure 2.

It is important to notice that, according to the model defined in Equation 2, the latent positions of each ego depend on both the latent positions of the other egos and the positions of the alters. In fact, conditional on the ego-ego relations, the latent positions of alters influences the positions of the egos by shortening or lengthen their distance in the latent space. It may happen that two unconnected egos tend to be conditionally close to each other in the latent space because they share a large number of common alters and, vice versa, two connected egos tend to be conditionally far from each other because they either do not share or share only a few common alters.

2.3 Variational inference

Several methods have been proposed to estimate model parameters and nodal latent positions. These methods include Monte Carlo algorithms from stationary distributions corresponding to the posterior distributions (Hoff et al., 2002; Handcock et al., 2007; Krivitsky et al., 2009; Raftery et al., 2012). Variational methods (Jordan et al., 1999) offer a fast approximate alternative inferential methodology for large data sets (Salter-Townshend and Murphy, 2013; Gollini and Murphy, 2015).

Due to the large size of the data we analyse in this paper, variational methods represent a pragmatic and effective choice. The target posterior distribution corresponding to the model defined in Equation 2 can be written as

$$
\begin{aligned}
p(\mathbf{Z}, \mathbf{W}, \alpha,\ \beta \mid \mathbf{X}, \mathbf{Y}) \propto\ & p(\mathbf{Y}, \mathbf{X} \mid \mathbf{Z}, \mathbf{W}, \beta, \alpha) \times \\
& p(\mathbf{Z})\ p(\mathbf{W})\ p(\beta)\ p(\alpha),
\end{aligned}
$$

where the distributions of $p(\mathbf{Z})$, $p(\mathbf{W})$, $p(\beta)$, and $p(\alpha)$ are defined in Section 2.1 and 2.2.

We propose the following variational approximation to the target distribution:

$$q(\mathbf{Z}, \mathbf{W}, \alpha, \beta \mid \mathbf{Y}, \mathbf{X}) = q(\alpha)\ q(\beta) \prod_{i=1}^{N} q(\mathbf{z}_i) \prod_{k=1}^{M} q(\mathbf{w}_k),$$

where $q(\alpha) = \mathcal{N}(\tilde{\xi}_\alpha, \tilde{\psi}_\alpha^2)$, $q(\beta) = \mathcal{N}(\tilde{\xi}_\beta, \tilde{\psi}_\beta^2)$, $q(\mathbf{z}_i) = \mathcal{N}(\tilde{\mathbf{z}}_i, \tilde{\mathbf{\Sigma}}_z)$ and $q(\mathbf{w}_k) = \mathcal{N}(\tilde{\mathbf{w}}_k, \tilde{\mathbf{\Sigma}}_w)$.

An expectation-maximisation (EM) algorithm used to carry out parameter inference at each $(t+1)$ iteration consists of the following steps:

- E-Step:
  - Estimate $\tilde{\mathbf{z}}_i^{(t+1)}$ and $\tilde{\mathbf{\Sigma}}_z^{(t+1)}$ by evaluating:

$$\mathcal{Q}(\Theta_\alpha, \Theta_\beta; \Theta_\alpha^{(t)}, \Theta_\beta^{(t)}) = \mathrm{KL}[q(\mathbf{Z}, \alpha, \mathbf{W}, \beta \mid \mathbf{Y}, \mathbf{X}) \mid\mid p(\mathbf{Z}, \alpha, \mathbf{W}, \beta \mid \mathbf{Y}, \mathbf{X})],$$

    where $\mathrm{KL}(\cdot)$ is the Kullback–Leibler divergence measure and $\Theta_\alpha = (\tilde{\xi}_\alpha, \tilde{\psi}_\alpha^2)$, $\Theta_\beta = (\tilde{\xi}_\beta, \tilde{\psi}_\beta^2)$.
  - Estimate $\tilde{\mathbf{w}}_k^{(t+1)}$ and $\tilde{\mathbf{\Sigma}}_w^{(t+1)}$ by evaluating:

$$\mathcal{Q}(\Theta_\beta; \Theta_\beta^{(t)}) = \mathrm{KL}[q(\mathbf{Z}, \alpha, \mathbf{W}, \beta|\mathbf{Y}, \mathbf{X}) \mid\mid p(\mathbf{Z}, \alpha, \mathbf{W}, \beta|\mathbf{Y}, \mathbf{X})].$$

- M-Step:
  - Estimate $\tilde{\xi}_\alpha$ and $\tilde{\psi}_\alpha^2$ by evaluating:

$$\Theta_\alpha^{(t+1)} = \mathrm{argmax}\ \mathcal{Q}(\Theta_\alpha; \Theta_\alpha^{(t)}),$$

  - Estimate $\tilde{\xi}_\beta$ and $\tilde{\psi}_\beta^2$ by evaluating:

$$\Theta_\beta^{(t+1)} = \mathrm{argmax}\ \mathcal{Q}(\Theta_\beta; \Theta_\beta^{(t)}).$$

To minimise the risk of estimating local maxima, the algorithm needs to be run several times from different starting points. The solutions with the lowest expected log-likelihood is selected. Further mathematical details of the variational procedure are provided in the Appendix. Next, we apply our model to a dataset of wiretapped phone conversations exchanged among human smugglers.

3 Data collection and structure

On the 3rd of October 2013, a boat carrying 518 migrants capsized within sight of the island of Lampedusa, the southernmost Italian territory, claiming the life of 366 people on board (Nelson, 2014). The Italian authorities responded by launching an extensive police investigation. For the first time, the 'elite' Anti-mafia Prosecutor's Office in Palermo was tasked with investigating smuggling operations. Quickly, the experienced team of police officers was able to identify and then wiretap the phone lines of various individuals involved in the fatal journey. As it turned out, the individuals under surveillance were active in smuggling migrants across the Mediterranean sea along the so-called 'Central Mediterranean Route' (i.e., from Libya into Italy and, to a much lesser extent, Malta). This is one of the main smuggling routes into Europe with 693,731 illegal border crossings registered between 2013 and 2018, consistently accounting for half or more of all illegal entries into Europe (Frontex, 2019).

For this paper, we have acquired and analysed the complete set of phone records wiretapped among 28 smugglers active in the smuggling of migrants along the Central Mediterranean route. All the wiretapped smugglers were based in Italy at the time of the investigation. The phone records span from December 2013 to October 2014. This is a unique dataset that has never been used before and includes meta-data on 21,555 phone conversations wiretapped. During this period, the 28 smugglers under surveillance have been in contact with 15,791 individuals not under surveillance. To filter out the noise related to occasional contacts, we have restricted our analysis to alters (individuals not under surveillance) with degree greater than one. This will leave us with 28 egos and 2,687 alters.

We use this evidence to get a glimpse into the structure of interactions underpinning the market for human smuggling. Who are the central players in the market? Are there emerging clusters based on close interactions among egos /emphand shared alters? Who are the actors that display greater mutual criminal distance in their operations? And are there actors that possess a structurally equivalent profile? Granted, our 28 market players are just a slice of a much larger market; however, our evidence does offer some insights into the real-world behaviour of smugglers using a high-frequency data source (i.e., phone conversations wiretapped: on the use of this source in studying organised crime, see Campana and Varese (2013) and Campana and Varese (2020)). Our approach can be applied to much larger datasets, if available, to study a larger slice of this market and/or other illegal markets.

Formally, we treat the 28 smugglers as interdependent ego-networks. It should be noted here that all egos are smugglers (offenders) while alters can be anyone who has been in contact with them - both offenders and migrants (clients). Our evidence does not allow us to differentiate between offenders and migrants among the alters; however,

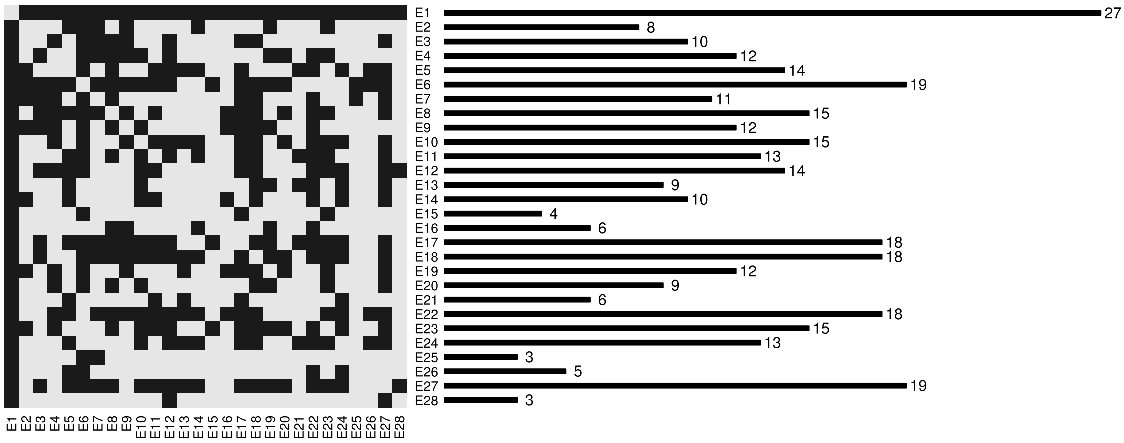

Figure 3: Adjacency Matrix **Y** and the ego degree for each ego.

it *does* allow us to model the full behaviour of the egos based on their interactions with both fellow offenders and migrants, i.e. both the supply-side and the demand-side of the market.

Figure 3 shows the adjacency matrix **Y** of the data set and the ego degree for each ego. Figure 4 shows the incidence matrix **X** and the alter degree for each ego. Figure 5 shows degree distribution for the alters.

4 Uncovering offenders' latent positions

We now move to explore the underlying structure of interactions in the market for human smuggling. We use our set of 28 smugglers (egos) to identify clusters of close interactions as well as what we can term 'criminal distancing'. We use the latent space as a representation of a criminal market, in this case the market for human smuggling. Before we do so, we remind the reader that interactions are based on phone conversations exchanged. An edge between any two actors is present if an interaction between them has been recorded. In this paper, we do not consider the direction of the call and the number of calls exchanged, hence we work with an undirected and unweighted graph. However, our modelling framework can be adapted for directed and weighted graphs.

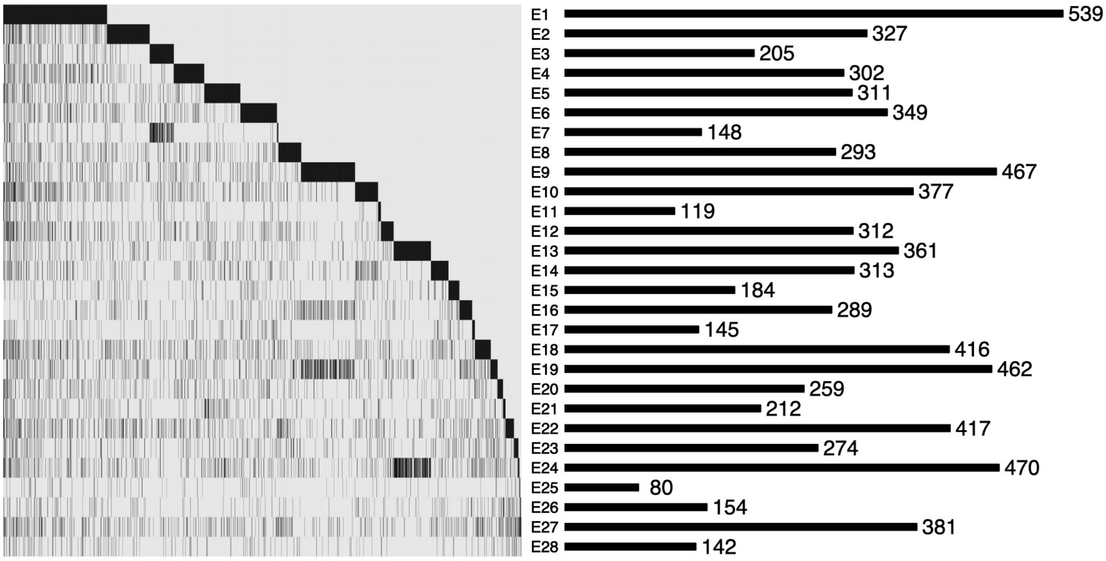

Figure 4: Incidence Matrix $\mathbf{X}$ and the alter degree for each ego.

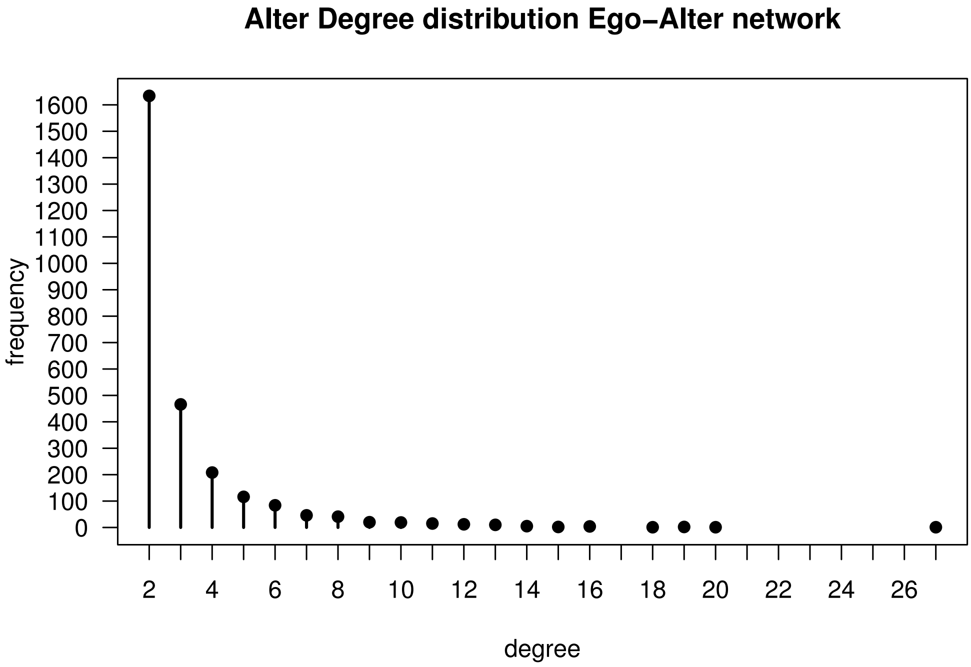

Figure 5: Degree distribution for the alters.

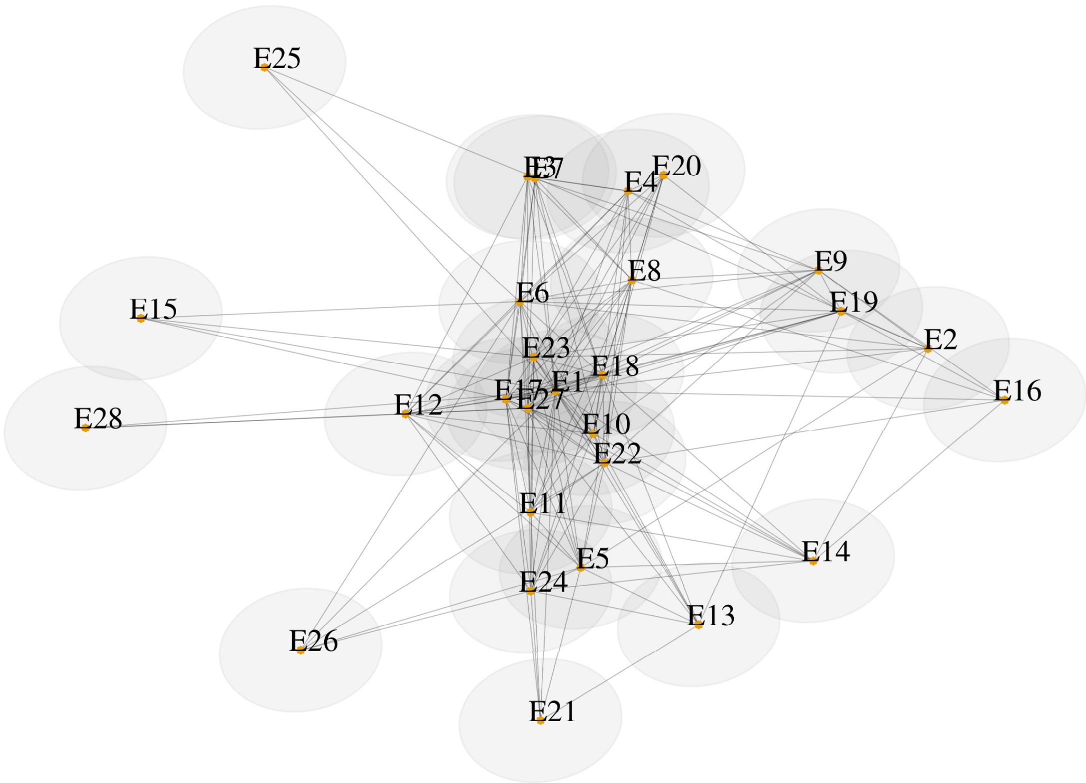

Figure 6: Estimate of latent positions and 95% credible intervals in grey on ego-ego network $\mathbf{Y}$.

4.1 Latent structure of the ego-ego network

We first analyse the 2-dimensional latent structure of the adjacency matrix $\mathbf{Y}$ of the ego-ego network without using the information about the incidence matrix $\mathbf{X}$ of the ego-alter relations (Equation 1). The code used to implement the methodology proposed in this paper is available in the lvm4net package (Gollini, 2020) for R (R Core Team, 2019). For the initialisation of the variational algorithm we used 10 random starting positions. We set the following fixed parameter values: $\sigma = 1$, $\xi_\alpha = 0$, and $\psi_\alpha^2 = 2$.

The estimated latent positions of the egos $\tilde{\mathbf{z}}_i$ are displayed in Figure 6 where the grey ellipses indicate the associated 95% credible intervals. The estimated posterior distribution of the density parameter α is a $\mathcal{N}(\tilde{\xi_\alpha} = 1.930, \tilde{\psi_\alpha}^2 = 0.003)$.

Figure 7 shows the graphical goodness of fit diagnostics for the estimated model. The procedure consists in comparing the distributions of network data simulated from the estimated variational posterior distributions to the observed data in terms of high-level network characteristics (Hunter et al., 2008). The plots suggest that the model is a reasonable fit to $\mathbf{Y}$ as the solid lines representing the observed network statistics lie within the 95% predictive network statistics intervals.

Figure 6 offers a first representation of the interactions underpinning the smuggling

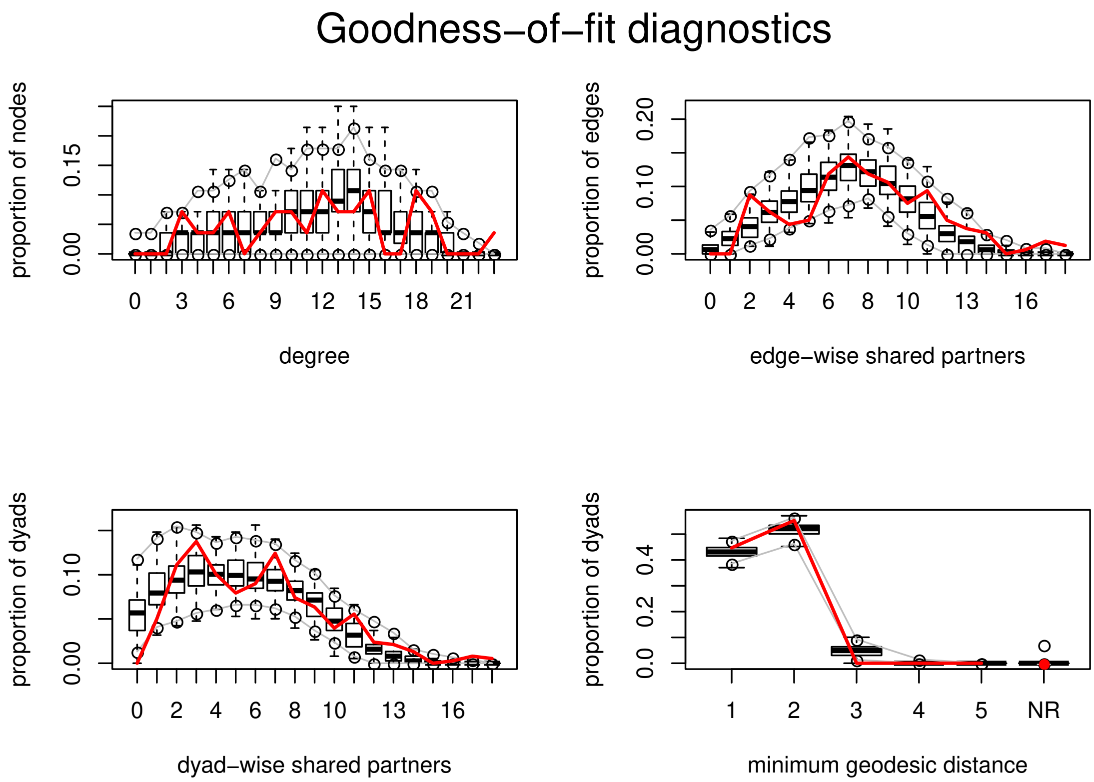

Figure 7: Graphical goodness of fit diagnostics for the latent space model on $\mathbf{Y}$. The solid lines represent the distribution of the observed network statistic distributions; the boxplots represent the simulated network statistic distributions.

market taking into considerations only the direct interactions among the 28 smugglers directly targeted. It neatly points to the centrality of E1 and a cluster of smugglers closely associated to him (all the smugglers in the dataset are male). It also shows the peripheral position of a number of other smugglers (e.g., E15, E25, E28, E21) as well as the presence of smugglers who cover the middle ground between the centre and the periphery, e.g. E12, E24, E8.

4.2 Latent structure of the interdependent ego-networks

We now include in our model the information about the interactions between egos (targeted smugglers) and alters (non-targeted individuals) to gain a complete picture of the behaviour of the 28 targeted smugglers. To do so, we estimate the model for interdependent ego-networks by including the relational information of the incidence matrix $\mathbf{X}$.

As for the previous section, we adopt a 2-dimensional latent space and the same initialisation specifications. We set the fixed parameters as following: $\sigma = 1$, $\xi_\alpha = 0$, $\psi_\alpha^2 = 1$, $\xi_\beta = 0$, and $\psi_\beta^2 = 2$.

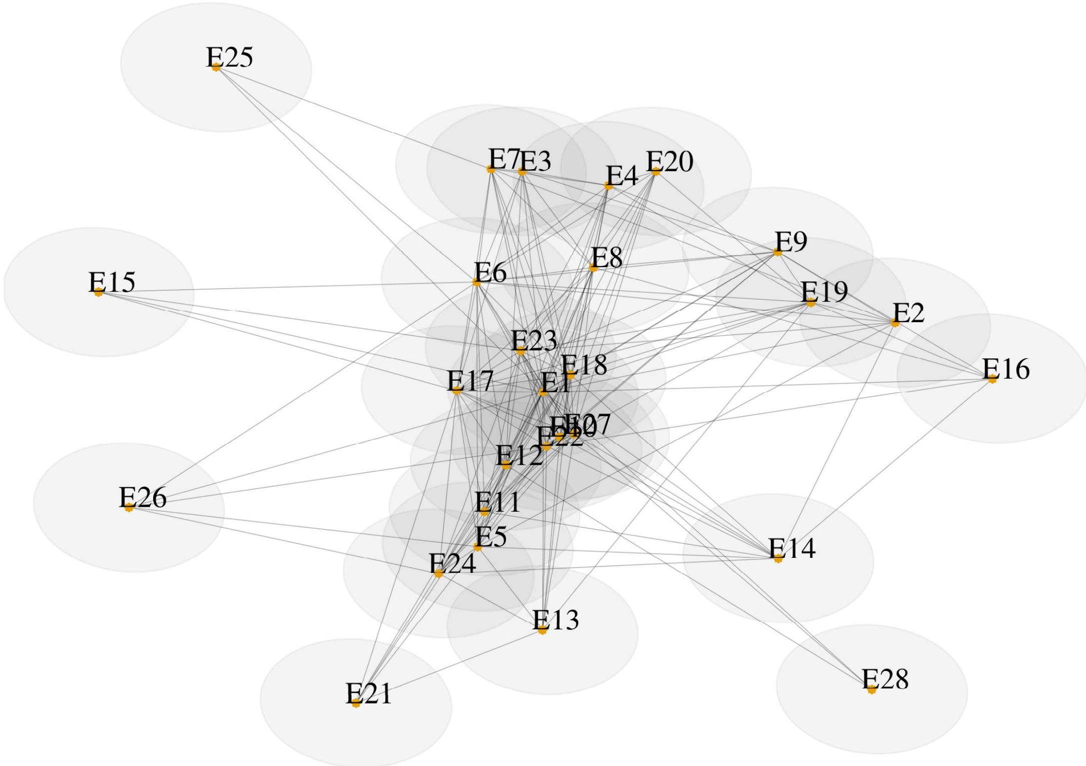

Figure 8: Estimate of latent positions and 95% credible intervals in grey on the interdependent ego network $(\mathbf{Y}, \mathbf{X})$.

The estimated latent positions of the egos $\tilde{\mathbf{z}}_i$ are displayed in Figure 8 where the grey ellipses indicate the associated 95% credible intervals. The estimated posterior distribution of the parameter α is $\mathcal{N}(\tilde{\xi}_\alpha = 1.754, \tilde{\psi}_\alpha^2 = 0.003)$. The estimated posterior distribution of the parameter β is $\mathcal{N}(\tilde{\xi}_\beta = -5.1974, \tilde{\psi}_\beta^2 = 0.0001)$.

Figure 8 offers the complete representation of the illicit market based on the full information on the behaviour of their (targeted) participants. We remind the reader that we use phone conversations wiretapped as a proxy for interactions. The analysis points to a number of findings. Firstly, the centrality of E1 and the presence of a dense cluster around him; these individuals may be his closest associates. Secondly, the presence of a second cluster far removed from the cluster around E1: this second cluster includes E9, E19, E2 and E16. There is a large 'criminal distance' between the two clusters, which may call for further investigation: are these two clusters in competition? Or are they fulfilling different tasks? The evidence we possess does not allow us to answer those questions. Further, there is a number of very peripheral actors, for instance E25, E15, E26 and E28. Finally, there is a number of players whose position is almost overlapping, e.g., E7/E3 and E4/E20. This suggests a high degree of equivalence from a market perspective, and thus the possibility that one could quickly replace the other if one of them is arrested (although this should be further investigated using qualitative evidence).

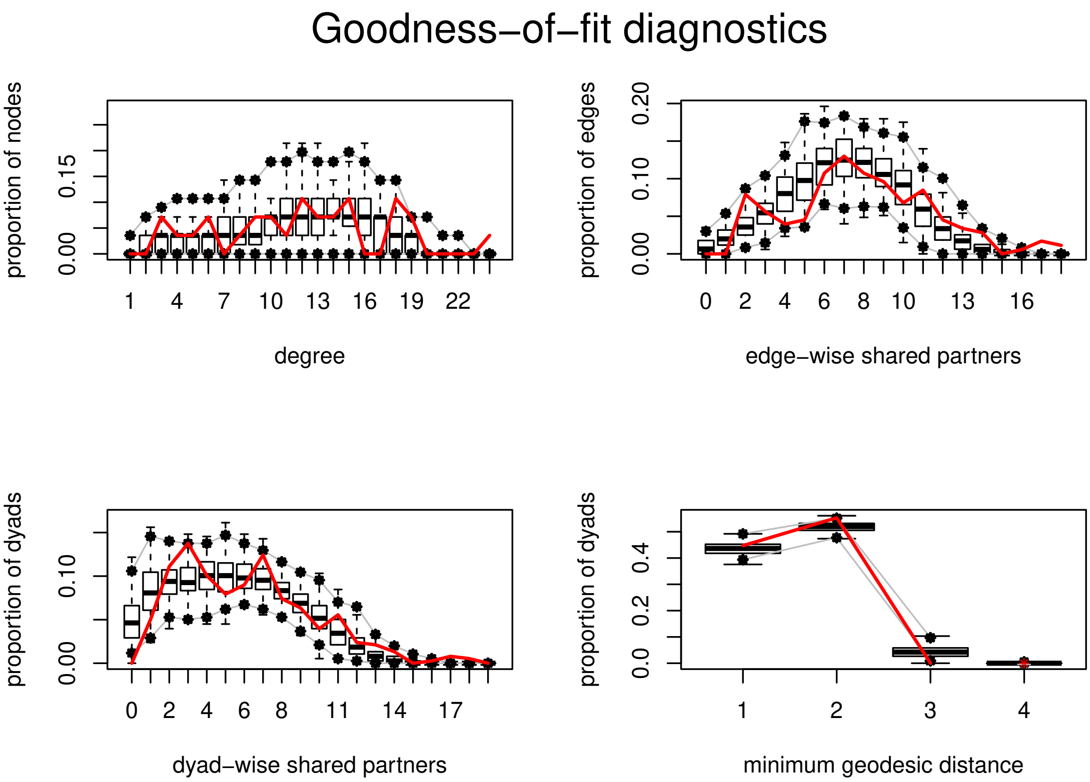

Figure 9: Graphical goodness of fit diagnostics for the latent space model on $(\mathbf{Y}, \mathbf{X})$.

When comparing Figure 8 to Figure 6, we can notice that most of the structure has remained unchanged. This means that the ego-alter relational structure is broadly reflecting the ego-ego relational structure. However, some changes did appear when using the full information available. Notably the position of E28, who was previously close to E15, is now showing a much higher criminal distance with E15 (and a full switch from the right-hand side to the left-hand side of the picture).

Figure 9 shows the graphical goodness of fit diagnostics for the estimated model for $(\mathbf{Y}, \mathbf{X})$. In this case the simulated network distributions are calculated on networks sampled from the estimated variational posterior distributions of the density parameters α, β and the ego and alters latent positions. The results confirm the good fit of the model to the ego-ego network data.

5 Estimating link probabilities

An interesting outcome of the model consists in inferring the link probabilities between any two egos i and j by using the expected posterior positions $\tilde{\mathbf{z}}_i$ and $\tilde{\mathbf{z}}_j$, and the expected density parameter value $\tilde{\xi_\alpha}$ (estimated from the latent space model for interconnected ego-networks):

$$\Pr(y_{ij} = 1 \mid \tilde{\mathbf{z}}_i, \tilde{\mathbf{z}}_j, \tilde{\xi_\alpha}) = \frac{\exp(\tilde{\xi_\alpha} - |\tilde{\mathbf{z}}_i - \tilde{\mathbf{z}}_j|^2)}{1 + \exp(\tilde{\xi_\alpha} - |\tilde{\mathbf{z}}_i - \tilde{\mathbf{z}}_j|^2)}, \tag{3}$$

The lower triangle of the matrix displayed in Figure 10 shows the estimated link probabilities for all the ego-ego network. The upper triangle represents the estimated link probabilities for empty dyads only. This allows us to identify those dyads (dark grey entries of the matrix) that, according to the estimated model, have similar connectivity patterns but are not connected to each other. These two egos can be seen as structurally equivalent with respect to both egos and alters. This is an important insight uncovered by the latent space model as it indicates the degree of equivalence of players in an illegal market based on their actual behaviour/interactions with both targeted and non-targeted market participants. This finding can have important implications for the disruption of illegal activities as it pinpoints individuals with a high degree of substitutability. (An alternative interpretation of this finding is that the two actors are in reality the same person; while we can rule out this possibility in our case given the evidence we possess, in other contexts the 'link-probability' analysis could help identify errors in the data, including the mis-identification of individuals.) The estimates of the link probability for each dyad in the network can further assist with predicting missing links. This could be done by using the mapping of the nodal latent distances together with additional information about the uncertainty about the actors' latent positions to train a classifier for predicting missing links.

6 Conclusions

Illegal markets are difficult to study; yet, they are part and parcel of our societies. The interactions underpinning such markets call for a network approach; however, scholars have found it difficult to collect primary data suitable for quantitative analysis, and thus have increasingly relied on evidence collected by the police. While this source can be very fruitful, it also poses challenges, as researchers have often no input in designing the data collection strategy. Therefore, they need to work within the constraints posed by the secondary nature of such evidence.

In this paper, we looked specifically at the effect of the sampling strategy adopted by law enforcement agencies on the data structure. We started from the observation that the targeted nature of police evidence creates two sets of actors: targeted individuals and non-targeted ones. We have then developed a latent space model for interdependent ego-networks purposely created to study interactions among offenders that (a) leverages on the full information available and (b) mirrors the specificity of the data collection strategy. We have suggested that this approach is suitable to model data directly stemming from police

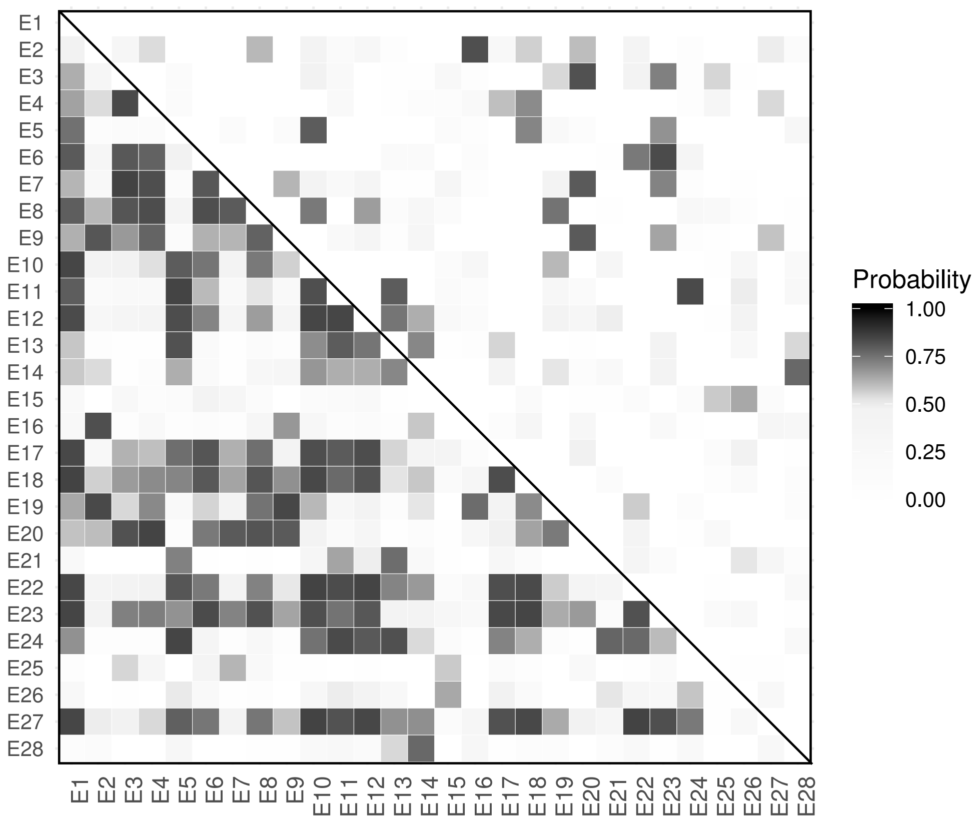

Figure 10: Latent space model on $(\mathbf{Y}, \mathbf{X})$: the lower triangle shows the estimated link probabilities for each ego dyad; the upper triangle shows the estimated link probabilities for empty ego dyads.

investigations as well as data extracted from police records using a targeted extraction approach.

We have posited that our tailored model can be fruitfully used to study interactions among offenders – and, more generally, the structure of illegal markets. By modelling a market as a latent space, researchers can identify central actors, clusters of close interactions ("criminal proximity") as well as gauging the reverse behaviour, which we have termed "criminal distancing". The model can also identify individuals who possess structurally equivalent market profiles. Our model is general as it can be applied to the study of any illegal market or, indeed, any type of interaction among offenders. Specific research questions will depend on the content of the evidence available; our approach only assumes that there are two sets of individuals: those who have been targeted and those who are known only by virtue of being connected to the targeted ones. Furthermore, our approach is based on a variational estimation procedure, which makes it very suitable for large networks like, for instance, criminal networks.

In this paper, we have applied our model to explore the underlying structure of interactions among 28 human smugglers (egos) using evidence from a high-frequency data source: phone conversations wiretapped by the police. We leveraged on the information about 21,555 phone calls recorded by the police among the 28 egos as well as between such egos and 2,686 alters. In the slice of the market for migrant smuggling under scrutiny, we have identified a central player and a cluster of individuals closely associated to him; we have also identified a second cluster far removed from the former suggesting either labour specialisation or competition. The analysis also uncovered the presence of peripheral actors, i.e., individuals showing a high criminal distance from any other actor in the market, including the two clusters. Further, we relied on the estimated link probabilities to study the degree of equivalence of actors in an illegal market. Our analysis takes into consideration the actual behaviour/interactions of egos with both targeted and non-targeted individuals. We believe this analysis can have relevant policy implications as it pinpoints individuals with a high degree of substitutability if arrested.

Future developments might expand the model to include actors' attributes, such as the specific task(s) carried out in the illegal market and their socio-demographic characteristics. The inclusion of actors' attributes, for instance in the form of specific tasks or the place where an actor is based, can help to better understand the function of the clusters identified as well as the reasons why certain actors display a large criminal distance or a close criminal proximity. The latent space framework presented in this paper can be easily extended to handle both ego and alter nodal/dyadic covariate information by specifying exogenous statistics with associated parameters capturing effects such as homophily, community structure, and heterogeneity of actors' characteristics (Krivitsky et al., 2009).

Acknowledgements

We are very grateful to Gery Ferrara, prosecutor with the Palermo Anti-mafia Office, for his invaluable assistance with the data collection, and to Carmine Mosca from the Italian Police for preparing the data. They have been extremely helpful and generous with their time. We are also grateful to the Editor and to the anonymous Reviewers for their helpful comments on a previous version of the paper.

References

Atkinson, R. and Flint, J. (2001), "Accessing hidden and hard-to-reach populations: Snowball research strategies," *Social Research Update*, 33, 1–4.

Berlusconi, G., Aziani, A., and Giommoni, L. (2017), "The determinants of heroin flows in Europe: A latent space approach," *Social Networks*, 51, 104–117.

Bright, D., Koskinen, J., and Malm, A. (2018), "Illicit network dynamics: The formation and evolution of a drug trafficking network," *Journal of Quantitative Criminology*, 1–22.

Bright, D. A., Hughes, C. E., and Chalmers, J. (2012), "Illuminating dark networks: A social network analysis of an Australian drug trafficking syndicate," *Crime, Law and Social Change*, 57, 151–176.

Calderoni, F. (2014), "Social network analysis of organized criminal groups," *Encyclopedia of Criminology and Criminal Justice*, 4972–4981.

Calderoni, F., Brunetto, D., and Piccardi, C. (2017), "Communities in criminal networks: A case study," *Social Networks*, 48, 116–125.

Campana, P. (2011), "Eavesdropping on the Mob: The functional diversification of Mafia activities across territories," *European Journal of Criminology*, 8, 213–228.

— (2016), "The structure of human trafficking: Lifting the bonnet on a Nigerian transnational network," *British Journal of Criminology*, 56, 68–86.

— (2018), "Out of Africa: The organization of migrant smuggling across the Mediterranean," *European Journal of Criminology*, 15, 481–502.

Campana, P. and Varese, F. (2012), "Listening to the wire: Criteria and techniques for the quantitative analysis of phone intercepts," *Trends in Organized Crime*, 15, 13–30.

— (2013), "Cooperation in criminal organizations: Kinship and violence as credible commitments," *Rationality and Society*, 25, 263–289.

— (2018), "Organized crime in the United Kingdom: Illegal governance of markets and communities," *The British Journal of Criminology*, 58, 1381–1400.

— (2020), "Studying organized crime networks: Data sources, boundaries and the limits of structural measures," *Social Networks*.

Diviák, T. (2019), "Key aspects of covert networks data collection: Problems, challenges, and opportunities," *Social Networks*.

Faust, K. and Tita, G. E. (2019), "Social networks and crime: Pitfalls and promises for advancing the field," *Annual Review of Criminology*, 2, 99–122.

Frontex (2019), *Risk Analysis for 2019*, Warsaw: Frontex.

Gollini, I. (2020), *lvm4net: Latent Variable Models for Networks.*, R package version 0.4.

Gollini, I. and Murphy, T. B. (2015), "Joint Modelling of Multiple Network Views," *Journal of Computational and Graphical Statistics*, 25, 246 – 265.

Grund, T. U. and Densley, J. A. (2012), "Ethnic heterogeneity in the activity and structure of a Black street gang," *European Journal of Criminology*, 9, 388–406.

Handcock, M. S., Raftery, A. E., and Tantrum, J. M. (2007), "Model-based clustering for social networks," *Journal Of The Royal Statistical Society Series A*, 170, 301–354.

Heckathorn, D. D. and Cameron, C. J. (2017), "Network sampling: From snowball and multiplicity to respondent-driven sampling," *Annual review of sociology*, 43, 101–119.

Hoff, P. D. (2005), "Bilinear mixed-effects models for dyadic data," *Journal of the American Statistical Association*, 100, 286–295.

— (2009), "Multiplicative latent factor models for description and prediction of social networks," *Computational and Mathematical Organization Theory*, 15, 261.

Hoff, P. D., Raftery, A. E., and Handcock, M. S. (2002), "Latent Space Approaches to Social Network Analysis," *Journal of the American Statistical Association*, 97, 1090–1098.

Hunter, D. R., Goodreau, S. M., and Handcock, M. S. (2008), "Goodness of Fit of Social Network Models," *Journal of the American Statistical Association*, 103, 248–258.

Jordan, M. I., Ghahramani, Z., Jaakkola, T. S., and Saul, L. K. (1999), "An Introduction to Variational Methods for Graphical Models," *Machine Learning*, 37, 183–233.

Krivitsky, P. N., Handcock, M. S., Raftery, A. E., and Hoff, P. D. (2009), "Representing degree distributions, clustering, and homophily in social networks with latent cluster random effects models," *Social networks*, 31, 204–213.

Malm, A. and Bichler, G. (2011), "Networks of Collaborating Criminals: Assessing the Structural Vulnerability of Drug Markets:," *Journal of Research in Crime and Delinquency*, 48, 271–297.

Malm, A. E., Kinney, J. B., and Pollard, N. R. (2008), "Social Network and Distance Correlates of Criminal Associates Involved in Illicit Drug Production," *Security Journal*, 21, 77–94.

McGloin, J. M. (2007), "Organizational Structure of Street Gangs in Newark, New Jersey: A Network Analysis Methodology," *Journal of Gang Research*, 15, 1–34.

McGloin, J. M. and Piquero, A. R. (2010), "On the Relationship between Co-Offending Network Redundancy and Offending Versatility:," *Journal of Research in Crime and Delinquency*, 47, 63–90.

Morselli, C. (2009), *Inside Criminal Networks*, Springer.

Morselli, C. (2013), *Crime and Networks*, Routledge.

Natarajan, M. (2000), "Understanding the Structure of a Drug Trafficking Organization: A Conversational Analysis," in *Illegal Drug Markets: from Research to Prevention Policy*, Monsey, NY: Criminal Justice Press, pp. 273–298.

Nelson, Z. (2014), "Lampedusa boat tragedy: A survivor's story," *The Guardian*.

Ouellet, M., Bouchard, M., and Charette, Y. (2019), "One gang dies, another gains? The network dynamics of criminal group persistence," *Criminology*, 57, 5–33.

Papachristos, A. V. (2009), "Murder by structure: dominance relations and the social structure of gang homicide." *American Journal of Sociology*, 115, 74–128.

Papachristos, A. V., Braga, A. A., and Hureau, D. M. (2012), "Social Networks and the Risk of Gunshot Injury," *Journal of Urban Health-bulletin of The New York Academy of Medicine*, 89, 992–1003.

Papachristos, A. V., Wildeman, C., and Roberto, E. (2015), "Tragic, but not random: The social contagion of nonfatal gunshot injuries," *Social Science & Medicine*, 125, 139–150.

R Core Team (2019), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.

Raftery, A. E., Niu, X., Hoff, P. D., and Yeung, K. Y. (2012), "Fast inference for the latent space network model using a case-control approximate likelihood," *Journal of Computational and Graphical Statistics*, 21, 901–919.

Rastelli, R., Friel, N., and Raftery, A. E. (2016), "Properties of latent variable network models," *Network Science*, 4, 407–432.

Salter-Townshend, M. and Murphy, T. B. (2013), "Variational Bayesian inference for the latent position cluster model for network data," *Computational Statistics & Data Analysis*, 57, 661–671.

Schaefer, D. R. (2012), "Youth co-offending networks: An investigation of social and spatial effects," *Social Networks*, 34, 141–149.

Smith, C. M. and Papachristos, A. V. (2016), "Trust thy crooked neighbor: multiplexity in Chicago organized crime networks," *American Sociological Review*, 81, 644–667.

Varese, F. (2013), "The Structure and the Content of Criminal Connections: The Russian Mafia in Italy," *European Sociological Review*, 29, 899–909.

Varese, F. (2013), "The structure and the content of criminal connections: The Russian Mafia in Italy," *European sociological review*, 29, 899–909.

Appendix

The ego-ego network latent space model is defined as:

$$p(\mathbf{Y}|\mathbf{Z}, \alpha) = \prod_{i \neq j}^{N} p(y_{ij}|\mathbf{z}_i, \mathbf{z}_j, \alpha) = \prod_{i \neq j}^{N} \frac{\exp(\alpha - |\mathbf{z}_i - \mathbf{z}_j|^2)^{y_{ij}}}{1 + \exp(\alpha - |\mathbf{z}_i - \mathbf{z}_j|^2)},$$

where for ease of notation $\prod_{i \neq j}^{N}$ is equivalent to $\prod_{i=1}^{N} \prod_{j=1, j \neq i}^{N}$.

We assume the following distributions for the model unknowns, where $p(\alpha) = \mathcal{N}(\xi_\alpha, \psi_\alpha^2)$, $p(\mathbf{z}_i) \overset{iid}{=} \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_D)$ and $\sigma^2, \xi_\alpha, \psi_\alpha^2$ are fixed parameters, and the squared Euclidean distance between ego i and ego j is $|\mathbf{z}_i - \mathbf{z}_j|^2 = (\mathbf{z}_i - \mathbf{z}_j)^\top (\mathbf{z}_i - \mathbf{z}_j) = \sum_{d=1}^{D} (z_{id} - z_{jd})^2$.

The ego-alter network latent space model is defined as:

$$p(\mathbf{X}|\mathbf{Z}, \mathbf{W}, \beta) = \prod_{i=1}^{N} \prod_{k=1}^{M} p(x_{ik}|\mathbf{z}_i, \mathbf{w}_k, \beta) = \prod_{i=1}^{N} \prod_{k=1}^{M} \frac{\exp(\beta - |\mathbf{z}_i - \mathbf{w}_k|^2)^{y_{ij}}}{1 + \exp(\beta - |\mathbf{z}_i - \mathbf{w}_k|^2)}.$$

We assume the following distributions for the model unknowns, where $p(\beta) = \mathcal{N}(\xi_\beta, \psi_\beta^2)$, $p(\mathbf{w}_k) \overset{iid}{=} \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_D)$ and $\sigma^2, \xi_\beta, \psi_\beta^2$ are fixed parameters, and the squared Euclidean distance between ego i and alter k is $|\mathbf{z}_i - \mathbf{w}_k|^2 = (\mathbf{z}_i - \mathbf{w}_k)^\top (\mathbf{z}_i - \mathbf{w}_k) = \sum_{d=1}^{D} (z_{id} - w_{kd})^2$.

The posterior probability is of the unknown $(\mathbf{Z}, \alpha)$ is of the form:

$$p(\mathbf{Z}, \mathbf{W}, \alpha, \beta|\mathbf{Y}, \mathbf{X}) \propto p(\mathbf{Y}|\mathbf{Z}, \alpha) p(\alpha) \prod_{i=1}^{N} p(\mathbf{z}_i) \times p(\mathbf{X}|\mathbf{Z}, \mathbf{W}, \beta) p(\beta) \prod_{k=1}^{M} p(\mathbf{w}_k).$$

We define the variational posterior $q(\mathbf{Z}, \mathbf{W}, \alpha, \beta|\mathbf{Y}, \mathbf{X})$ introducing the variational parameters $\tilde{\Theta}_\mathbf{z} = (\tilde{\xi}_\alpha, \tilde{\psi}_\alpha^2)$, $\tilde{\mathbf{z}}_i$, $\tilde{\mathbf{\Sigma}}_\mathbf{z}$, $\tilde{\Theta}_\mathbf{w} = (\tilde{\xi}_\beta, \tilde{\psi}_\beta^2)$, $\tilde{\mathbf{w}}_k$, $\tilde{\mathbf{\Sigma}}_\mathbf{w}$:

$$q(\mathbf{Z}, \mathbf{W}, \alpha, \beta|\mathbf{Y}, \mathbf{X}) = q(\alpha) q(\beta) \prod_{i=1}^{N} q(\mathbf{z}_i) \prod_{k=1}^{M} q(\mathbf{w}_k),$$

where $q(\alpha) = \mathcal{N}(\tilde{\xi}_\alpha, \tilde{\psi}_\alpha^2)$, $q(\mathbf{z}_i) = \mathcal{N}(\tilde{\mathbf{z}}_i, \tilde{\mathbf{\Sigma}}_\mathbf{z})$, $q(\beta) = \mathcal{N}(\tilde{\xi}_\beta, \tilde{\psi}_\beta^2)$ and $q(\mathbf{w}_k) = \mathcal{N}(\tilde{\mathbf{w}}_k, \tilde{\mathbf{\Sigma}}_\mathbf{w})$.

6.1 Kullback Leibler divergence

$$
\begin{aligned}
&\mathrm{KL}[q(\mathbf{Z},\mathbf{W},\alpha,\beta|\mathbf{Y},\mathbf{X})||p(\mathbf{Z},\mathbf{W},\alpha,\beta|\mathbf{Y},\mathbf{X})] \\
&\quad= \mathrm{KL}[q(\alpha)||p(\alpha)] + \mathrm{KL}[q(\beta)||p(\beta)] + \sum_{i=1}^{N}\mathrm{KL}[q(\mathbf{z}_i)||p(\mathbf{z}_i)] + \sum_{k=1}^{M}\mathrm{KL}[q(\mathbf{w}_k)||p(\mathbf{w}_k)] \\
&\qquad - \mathbb{E}_{q(\mathbf{Z},\mathbf{W},\alpha,\beta|\mathbf{Y},\mathbf{X})}[\log(p(\mathbf{Y}|\mathbf{Z},\alpha))] - \mathbb{E}_{q(\mathbf{Z},\mathbf{W},\alpha,\beta|\mathbf{Y},\mathbf{X})}[\log(p(\mathbf{X}|\mathbf{Z},\mathbf{W},\beta))] \\
&\quad= \frac{1}{2}\left(\frac{\tilde{\psi}_\alpha^2}{\psi_\alpha^2} - \log\frac{\tilde{\psi}_\alpha^2}{\psi_\alpha^2} + \frac{(\tilde{\xi}_\alpha-\xi_\alpha)^2}{\psi_\alpha^2} + \frac{\tilde{\psi}_\beta^2}{\psi_\beta^2} - \log\frac{\tilde{\psi}_\beta^2}{\psi_\beta^2} + \frac{(\tilde{\xi}_\beta-\xi_\beta)^2}{\psi_\beta^2}\right. \\
&\qquad\qquad \left. + ND\log(\sigma^2) - N\log(\det(\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}})) + MD\log(\sigma^2) - M\log(\det(\tilde{\boldsymbol{\Sigma}}_{\mathbf{w}}))\right) \\
&\qquad + \frac{N}{2\sigma^2}\mathrm{tr}(\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}}) + \frac{\sum_{i=1}^{N}\tilde{\mathbf{z}}_i^\top\tilde{\mathbf{z}}_i}{2\sigma^2} + \frac{M}{2\sigma^2}\mathrm{tr}(\tilde{\boldsymbol{\Sigma}}_{\mathbf{w}}) + \frac{\sum_{k=1}^{M}\tilde{\mathbf{w}}_k^\top\tilde{\mathbf{w}}_k}{2\sigma^2} - \frac{1+(N+M)D}{2} \\
&\qquad - \mathbb{E}_{q(\mathbf{Z},\mathbf{W},\alpha,\beta|\mathbf{Y},\mathbf{X})}[\log(p(\mathbf{Y}|\mathbf{Z},\alpha))] - \mathbb{E}_{q(\mathbf{Z},\mathbf{W},\alpha,\beta|\mathbf{Y},\mathbf{X})}[\log(p(\mathbf{X}|\mathbf{Z},\mathbf{W},\beta))] \\
&\quad\leq \frac{1}{2}\left(\frac{\tilde{\psi}_\alpha^2}{\psi_\alpha^2} - \log\frac{\tilde{\psi}_\alpha^2}{\psi_\alpha^2} + \frac{(\tilde{\xi}_\alpha-\xi_\alpha)^2}{\psi_\alpha^2} + \frac{\tilde{\psi}_\beta^2}{\psi_\beta^2} - \log\frac{\tilde{\psi}_\beta^2}{\psi_\beta^2} + \frac{(\tilde{\xi}_\beta-\xi_\beta)^2}{\psi_\beta^2}\right. \\
&\qquad\qquad \left. + ND\log(\sigma^2) - N\log(\det(\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}})) + MD\log(\sigma^2) - M\log(\det(\tilde{\boldsymbol{\Sigma}}_{\mathbf{w}}))\right) \\
&\qquad + \frac{N}{2\sigma^2}\mathrm{tr}(\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}}) + \frac{\sum_{i=1}^{N}\tilde{\mathbf{z}}_i^\top\tilde{\mathbf{z}}_i}{2\sigma^2} + \frac{M}{2\sigma^2}\mathrm{tr}(\tilde{\boldsymbol{\Sigma}}_{\mathbf{w}}) + \frac{\sum_{k=1}^{M}\tilde{\mathbf{w}}_k^\top\tilde{\mathbf{w}}_k}{2\sigma^2} - \frac{1+(N+M)D}{2} \\
&\qquad + \sum_{i\neq j}^{N} y_{ij}(-\tilde{\xi}_\alpha + 2\mathrm{tr}(\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}}) + |\tilde{\mathbf{z}}_i - \tilde{\mathbf{z}}_j|^2) \\
&\qquad\qquad + \log\left(1 + \frac{\exp\left(\tilde{\xi}_\alpha + \frac{1}{2}\tilde{\psi}_\alpha^2\right)}{\det(\mathbf{I}+4\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}})^{\frac{1}{2}}}\exp\left(-(\tilde{\mathbf{z}}_i-\tilde{\mathbf{z}}_j)^\top(\mathbf{I}+4\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}})^{-1}(\tilde{\mathbf{z}}_i-\tilde{\mathbf{z}}_j)\right)\right) \\
&\qquad + \sum_{i=1}^{N}\sum_{k=1}^{M} x_{ik}(-\tilde{\xi}_\beta + \mathrm{tr}(\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}} + \tilde{\boldsymbol{\Sigma}}_{\mathbf{w}}) + |\tilde{\mathbf{z}}_i - \tilde{\mathbf{w}}_k|^2) \\
&\qquad\qquad\quad + \log\left(1 + \frac{\exp\left(\tilde{\xi}_\beta + \frac{1}{2}\tilde{\psi}_\beta^2\right)}{\det(\mathbf{I}+2(\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}}+\tilde{\boldsymbol{\Sigma}}_{\mathbf{w}}))^{\frac{1}{2}}}\exp\left(-(\tilde{\mathbf{z}}_i-\tilde{\mathbf{w}}_k)^\top(\mathbf{I}+2(\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}}+\tilde{\boldsymbol{\Sigma}}_{\mathbf{w}}))^{-1}(\tilde{\mathbf{z}}_i-\tilde{\mathbf{w}}_k)\right)\right),
\end{aligned}
$$

where $\mathbb{E}_{q(\mathbf{Z},\alpha|\mathbf{Y})}[\log(p(\mathbf{Y}|\mathbf{Z},\alpha))]$ is approximated using the Jensen's inequality:

$$
\begin{aligned}
\mathbb{E}_q[\log(p(\mathbf{Y}|\mathbf{Z},\alpha))] &= \sum_{i\neq j}^{N} y_{ij}\mathbb{E}_{q(\mathbf{Z},\alpha|\mathbf{Y})}[\alpha - |\mathbf{z}_i - \mathbf{z}_j|^2] - \mathbb{E}_{q(\mathbf{Z},\alpha|\mathbf{Y})}[\log(1+\exp(\alpha - |\mathbf{z}_i - \mathbf{z}_j|^2))] \\
&\leq \sum_{i\neq j}^{N} y_{ij}(\mathbb{E}_{q(\mathbf{Z},\alpha|\mathbf{Y})}[\alpha - |\mathbf{z}_i - \mathbf{z}_j|^2]) - \log(1+\mathbb{E}_{q(\mathbf{Z},\alpha|\mathbf{Y})}[\exp(\alpha - |\mathbf{z}_i - \mathbf{z}_j|^2)]) \\
&= \sum_{i\neq j}^{N} y_{ij}(\tilde{\xi}_\alpha - 2\mathrm{tr}(\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}}) - |\tilde{\mathbf{z}}_i - \tilde{\mathbf{z}}_j|^2) \\
&\quad - \log\left(1 + \frac{\exp\left(\tilde{\xi}_\alpha + \frac{1}{2}\tilde{\psi}_\alpha^2\right)}{\det(\mathbf{I}+4\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}})^{\frac{1}{2}}}\exp\left(-(\tilde{\mathbf{z}}_i - \tilde{\mathbf{z}}_j)^\top(\mathbf{I}+4\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}})^{-1}(\tilde{\mathbf{z}}_i - \tilde{\mathbf{z}}_j)\right)\right).
\end{aligned}
$$

and $\mathbb{E}_{q(\mathbf{Z},\mathbf{W},\alpha,\beta|\mathbf{Y},\mathbf{X})}[\log(p(\mathbf{X}|\mathbf{Z},\mathbf{W},\beta))]$ is approximated using the Jensen's inequality:

$$
\begin{aligned}
\mathbb{E}_q[\log(p(\mathbf{X}|\mathbf{Z},\mathbf{W},\beta))] &= \sum_{i=1}^{N}\sum_{k=1}^{M} x_{ik}\mathbb{E}_q[\beta - |\mathbf{z}_i - \mathbf{w}_k|^2] - \mathbb{E}_q[\log(1+\exp(\beta - |\mathbf{z}_i - \mathbf{w}_k|^2))] \\
&\leq \sum_{i=1}^{N}\sum_{k=1}^{M} x_{ik}(\mathbb{E}_q[\beta - |\mathbf{z}_i - \mathbf{w}_k|^2]) - \log(1+\mathbb{E}_q[\exp(\beta - |\mathbf{z}_i - \mathbf{w}_k|^2)]) \\
&= \sum_{i=1}^{N}\sum_{k=1}^{M} x_{ik}(\tilde{\xi}_\beta - \mathrm{tr}(\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}} + \tilde{\boldsymbol{\Sigma}}_{\mathbf{w}}) - |\tilde{\mathbf{z}}_i - \tilde{\mathbf{w}}_k|^2) \\
&\quad - \log\left(1 + \frac{\exp\left(\tilde{\xi}_\beta + \frac{1}{2}\tilde{\psi}_\beta^2\right)}{\det(\mathbf{I}+2(\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}} + \tilde{\boldsymbol{\Sigma}}_{\mathbf{w}}))^{\frac{1}{2}}}\exp\left(-(\tilde{\mathbf{z}}_i - \tilde{\mathbf{w}}_k)^\top(\mathbf{I}+2(\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}} + \tilde{\boldsymbol{\Sigma}}_{\mathbf{w}}))^{-1}(\tilde{\mathbf{z}}_i - \tilde{\mathbf{w}}_k)\right)\right).
\end{aligned}
$$

6.2 Estimate $\tilde{\mathbf{z}}_i$

$$
\begin{aligned}
\mathrm{KL} \leq \mathrm{Const}_{\tilde{\mathbf{z}}_i} &+ \tilde{\mathbf{z}}_i^\top\tilde{\mathbf{z}}_i\left(\frac{1}{2\sigma^2} + \sum_{j\neq i}(y_{ji}+y_{ij}) + \sum_{k=1}^{M} x_{ik}\right) - 2\tilde{\mathbf{z}}_i^\top\left(\sum_{i\neq j}(y_{ji}+y_{ij})\tilde{\mathbf{z}}_j + \sum_{k=1}^{M} x_{ik}\tilde{\mathbf{w}}_k\right) \\
&+ 2\sum_{j\neq i}\log\left(1 + \frac{\exp\left(\tilde{\xi}_\alpha + \frac{1}{2}\tilde{\psi}_\alpha^2\right)}{\det(\mathbf{I}+4\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}})^{\frac{1}{2}}}\exp\left(-(\tilde{\mathbf{z}}_i - \tilde{\mathbf{z}}_j)^\top(\mathbf{I}+4\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}})^{-1}(\tilde{\mathbf{z}}_i - \tilde{\mathbf{z}}_j)\right)\right) \\
&+ \sum_{k=1}^{M}\log\left(1 + \frac{\exp\left(\tilde{\xi}_\beta + \frac{1}{2}\tilde{\psi}_\beta^2\right)}{\det(\mathbf{I}+2(\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}} + \tilde{\boldsymbol{\Sigma}}_{\mathbf{w}}))^{\frac{1}{2}}}\exp\left(-(\tilde{\mathbf{z}}_i - \tilde{\mathbf{w}}_k)^\top(\mathbf{I}+2(\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}} + \tilde{\boldsymbol{\Sigma}}_{\mathbf{w}}))^{-1}(\tilde{\mathbf{z}}_i - \tilde{\mathbf{w}}_k)\right)\right).
\end{aligned}
$$

Second-order Taylor series expansion approximation of

$$f(\tilde{\mathbf{z}}_i) = \sum_{j \neq i} \log \left(1 + \frac{\exp\left(\tilde{\xi}_\alpha + \frac{1}{2}\tilde{\psi}_\alpha^2\right)}{\det(\mathbf{I} + 4\tilde{\mathbf{\Sigma}}_\mathbf{z})^{\frac{1}{2}}} \exp\left(-(\tilde{\mathbf{z}}_i - \tilde{\mathbf{z}}_j)^\top (\mathbf{I} + 4\tilde{\mathbf{\Sigma}}_\mathbf{z})^{-1} (\tilde{\mathbf{z}}_i - \tilde{\mathbf{z}}_j)\right) \right)$$
$$+ \frac{1}{2} \sum_{k=1}^{M} \log \left(1 + \frac{\exp\left(\tilde{\xi}_\beta + \frac{1}{2}\tilde{\psi}_\beta^2\right)}{\det(\mathbf{I} + 2(\tilde{\mathbf{\Sigma}}_\mathbf{z} + \tilde{\mathbf{\Sigma}}_\mathbf{w}))^{\frac{1}{2}}} \exp\left(-(\tilde{\mathbf{z}}_i - \tilde{\mathbf{w}}_k)^\top (\mathbf{I} + 2(\tilde{\mathbf{\Sigma}}_\mathbf{z} + \tilde{\mathbf{\Sigma}}_\mathbf{w}))^{-1} (\tilde{\mathbf{z}}_i - \tilde{\mathbf{w}}_k)\right) \right).$$

Therefore,

$$f(\tilde{\mathbf{z}}_i) \approx f(\tilde{\mathbf{z}}_i^o) + (\tilde{\mathbf{z}}_i - \tilde{\mathbf{z}}_i^o)^\top G(\tilde{\mathbf{z}}_i^o) + \frac{1}{2}(\tilde{\mathbf{z}}_i - \tilde{\mathbf{z}}_i^o)^\top H(\tilde{\mathbf{z}}_i^o)(\tilde{\mathbf{z}}_i - \tilde{\mathbf{z}}_i^o).$$

Let's find the gradient G and the Hessian matrix H of f evaluated at $\tilde{\mathbf{z}}_i = \tilde{\mathbf{z}}_i^o$.

$$\begin{aligned}
G(\tilde{\mathbf{z}}_i^o) = & -2(\mathbf{I} + 4\tilde{\mathbf{\Sigma}}_\mathbf{z})^{-1} \sum_{j \neq i} (\tilde{\mathbf{z}}_i^o - \tilde{\mathbf{z}}_j) \\
& \times \left[1 + \frac{\det(\mathbf{I} + 4\tilde{\mathbf{\Sigma}}_\mathbf{z})^{\frac{1}{2}}}{\exp\left(\tilde{\xi}_\alpha + \frac{1}{2}\tilde{\psi}_\alpha^2\right)} \exp\left((\tilde{\mathbf{z}}_i^o - \tilde{\mathbf{z}}_j)^\top (\mathbf{I} + 4\tilde{\mathbf{\Sigma}}_\mathbf{z})^{-1} (\tilde{\mathbf{z}}_i^o - \tilde{\mathbf{z}}_j)\right) \right]^{-1} \\
& - (\mathbf{I} + 2(\tilde{\mathbf{\Sigma}}_\mathbf{z} + \tilde{\mathbf{\Sigma}}_\mathbf{w}))^{-1} \sum_{k=1}^{M} (\tilde{\mathbf{z}}_i^o - \tilde{\mathbf{w}}_k) \\
& \times \left[1 + \frac{\det(\mathbf{I} + 2(\tilde{\mathbf{\Sigma}}_\mathbf{z} + \tilde{\mathbf{\Sigma}}_\mathbf{w}))^{\frac{1}{2}}}{\exp\left(\tilde{\xi}_\beta + \frac{1}{2}\tilde{\psi}_\beta^2\right)} \exp\left((\tilde{\mathbf{z}}_i^o - \tilde{\mathbf{w}}_k)^\top (\mathbf{I} + 2(\tilde{\mathbf{\Sigma}}_\mathbf{z} + \tilde{\mathbf{\Sigma}}_\mathbf{w}))^{-1} (\tilde{\mathbf{z}}_i^o - \tilde{\mathbf{w}}_k)\right) \right]^{-1}.
\end{aligned}$$

$$
\begin{aligned}
H(\tilde{\mathbf{z}}_i^o) = & -2(\mathbf{I}+4\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}})^{-1}\sum_{j\neq i}\left[1+\frac{\det(\mathbf{I}+4\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}})^{\frac{1}{2}}}{\exp\left(\tilde{\xi}_\alpha+\frac{1}{2}\tilde{\psi}_\alpha^2\right)}\exp\left((\tilde{\mathbf{z}}_i^o-\tilde{\mathbf{z}}_j)^\top(\mathbf{I}+4\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}})^{-1}(\tilde{\mathbf{z}}_i^o-\tilde{\mathbf{z}}_j)\right)\right]^{-1} \\
& \times\left[\mathbf{I}-\frac{2(\tilde{\mathbf{z}}_i^o-\tilde{\mathbf{z}}_j)(\tilde{\mathbf{z}}_i^o-\tilde{\mathbf{z}}_j)^\top(\mathbf{I}+4\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}})^{-1}}{1+\frac{\exp\left(\tilde{\xi}_\alpha+\frac{1}{2}\tilde{\psi}_\alpha^2\right)}{\det(\mathbf{I}+4\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}})^{\frac{1}{2}}}\exp\left(-(\tilde{\mathbf{z}}_i^o-\tilde{\mathbf{z}}_j)^\top(\mathbf{I}+4\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}})^{-1}(\tilde{\mathbf{z}}_i^o-\tilde{\mathbf{z}}_j)\right)}\right] \\
& -(\mathbf{I}+2(\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}}+\tilde{\boldsymbol{\Sigma}}_{\mathbf{w}}))^{-1} \\
& \times\sum_{k=1}^{M}\left[1+\frac{\det(\mathbf{I}+2(\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}}+\tilde{\boldsymbol{\Sigma}}_{\mathbf{w}}))^{\frac{1}{2}}}{\exp\left(\tilde{\xi}_\beta+\frac{1}{2}\tilde{\psi}_\beta^2\right)}\exp\left((\tilde{\mathbf{z}}_i^o-\tilde{\mathbf{w}}_k)^\top(\mathbf{I}+2(\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}}+\tilde{\boldsymbol{\Sigma}}_{\mathbf{w}}))^{-1}(\tilde{\mathbf{z}}_i^o-\tilde{\mathbf{w}}_k)\right)\right]^{-1} \\
& \times\left[\mathbf{I}-\frac{2(\tilde{\mathbf{z}}_i^o-\tilde{\mathbf{w}}_k)(\tilde{\mathbf{z}}_i^o-\tilde{\mathbf{w}}_k)^\top(\mathbf{I}+2(\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}}+\tilde{\boldsymbol{\Sigma}}_{\mathbf{w}}))^{-1}}{1+\frac{\exp\left(\tilde{\xi}_\beta+\frac{1}{2}\tilde{\psi}_\beta^2\right)}{\det(\mathbf{I}+2(\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}}+\tilde{\boldsymbol{\Sigma}}_{\mathbf{w}}))^{\frac{1}{2}}}\exp\left(-(\tilde{\mathbf{z}}_i^o-\tilde{\mathbf{w}}_k)^\top(\mathbf{I}+2(\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}}+\tilde{\boldsymbol{\Sigma}}_{\mathbf{w}}))^{-1}(\tilde{\mathbf{z}}_i^o-\tilde{\mathbf{w}}_k)\right)}\right].
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\mathrm{KL} \approx\ & \tilde{\mathbf{z}}_i^\top\left[\left(\frac{1}{2\sigma^2}+\sum_{j\neq i}(y_{ji}+y_{ij})+\sum_{k=1}^{M}x_{ik}\right)\mathbf{I}+H(\tilde{\mathbf{z}}_i^o)\right]\tilde{\mathbf{z}}_i \\
& -2\tilde{\mathbf{z}}_i^\top\left[\sum_{j\neq i}(y_{ji}+y_{ij})\tilde{\mathbf{z}}_j+\sum_{k=1}^{M}x_{ik}\tilde{\mathbf{w}}_k-G(\tilde{\mathbf{z}}_i^o)+H(\tilde{\mathbf{z}}_i^o)\tilde{\mathbf{z}}_i^o\right]+\mathrm{Const}_{\tilde{\mathbf{z}}_i}.
\end{aligned}
$$

Set $\dfrac{\partial \mathrm{KL}}{\partial \tilde{\mathbf{z}}_i}=0:$

$$
\begin{aligned}
\tilde{\mathbf{z}}_i = & \left[\left(\frac{1}{2\sigma^2}+\sum_{j\neq i}(y_{ji}+y_{ij})+\sum_{k=1}^{M}x_{ik}\right)\mathbf{I}+H(\tilde{\mathbf{z}}_i^o)\right]^{-1} \\
& \times\left[\sum_{j\neq i}(y_{ji}+y_{ij})\tilde{\mathbf{z}}_j+\sum_{k=1}^{M}x_{ik}\tilde{\mathbf{w}}_k-G(\tilde{\mathbf{z}}_i^o)+H(\tilde{\mathbf{z}}_i^o)\tilde{\mathbf{z}}_i^o\right].
\end{aligned}
$$

6.3 Estimate $\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}}$

$$
\begin{aligned}
\mathrm{KL} \leq & \mathrm{Const}_{\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}}} + \mathrm{tr}(\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}}) \left(\frac{N}{2\sigma^2} + 2\sum_{i \neq j} y_{ij} + \sum_{i=1}^{N} \sum_{k=1}^{M} x_{ik} \right) - \frac{N}{2} \log(\det(\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}})) \\
& + \sum_{i \neq j} \log \left(1 + \frac{\exp\left(\tilde{\xi}_{\alpha} + \frac{1}{2}\tilde{\psi}_{\alpha}^2\right)}{\det(\mathbf{I} + 4\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}})^{\frac{1}{2}}} \exp\left(-(\tilde{\mathbf{z}}_i - \tilde{\mathbf{z}}_j)^\top (\mathbf{I} + 4\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}})^{-1} (\tilde{\mathbf{z}}_i - \tilde{\mathbf{z}}_j) \right) \right) \\
& + \sum_{i=1}^{N} \sum_{k=1}^{M} \log \left(1 + \frac{\exp\left(\tilde{\xi}_{\beta} + \frac{1}{2}\tilde{\psi}_{\beta}^2\right)}{\det(\mathbf{I} + 2(\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}} + \tilde{\boldsymbol{\Sigma}}_{\mathbf{w}}))^{\frac{1}{2}}} \exp\left(-(\tilde{\mathbf{z}}_i - \tilde{\mathbf{w}}_k)^\top (\mathbf{I} + 2(\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}} + \tilde{\boldsymbol{\Sigma}}_{\mathbf{w}}))^{-1} (\tilde{\mathbf{z}}_i - \tilde{\mathbf{w}}_k) \right) \right).
\end{aligned}
$$

First-order Taylor series expansion approximation of

$$
\begin{aligned}
f(\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}}) = & \sum_{i \neq j} \log \left(1 + \frac{\exp\left(\tilde{\xi}_{\alpha} + \frac{1}{2}\tilde{\psi}_{\alpha}^2\right)}{\det(\mathbf{I} + 4\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}}^{o})^{\frac{1}{2}}} \exp\left(-(\tilde{\mathbf{z}}_i - \tilde{\mathbf{z}}_j)^\top (\mathbf{I} + 4\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}}^{o})^{-1} (\tilde{\mathbf{z}}_i - \tilde{\mathbf{z}}_j) \right) \right) \\
& + \sum_{i=1}^{N} \sum_{k=1}^{M} \log \left(1 + \frac{\exp\left(\tilde{\xi}_{\beta} + \frac{1}{2}\tilde{\psi}_{\beta}^2\right)}{\det(\mathbf{I} + 2(\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}} + \tilde{\boldsymbol{\Sigma}}_{\mathbf{w}}))^{\frac{1}{2}}} \exp\left(-(\tilde{\mathbf{z}}_i - \tilde{\mathbf{w}}_k)^\top (\mathbf{I} + 2(\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}} + \tilde{\boldsymbol{\Sigma}}_{\mathbf{w}}))^{-1} (\tilde{\mathbf{z}}_i - \tilde{\mathbf{w}}_k) \right) \right).
\end{aligned}
$$

We have

$$
f(\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}}) \approx f(\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}}^{o}) + J(\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}}^{o})(\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}} - \tilde{\boldsymbol{\Sigma}}_{\mathbf{z}}^{o}),
$$

where J is the Jacobian matrix of f evaluated at $\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}} = \tilde{\boldsymbol{\Sigma}}_{\mathbf{z}}^{o}$.

$$
\begin{aligned}
J(\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}}^{o}) = & 4(\mathbf{I} + 4\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}}^{o})^{-1} \sum_{i \neq j} \left((\tilde{\mathbf{z}}_i - \tilde{\mathbf{z}}_j)(\tilde{\mathbf{z}}_i - \tilde{\mathbf{z}}_j)^\top (\mathbf{I} + 4\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}}^{o})^{-1} - \frac{1}{2}\mathbf{I} \right) \cdot \\
& \cdot \left[1 + \frac{\det(\mathbf{I} + 4\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}}^{o})^{\frac{1}{2}}}{\exp\left(\tilde{\xi}_{\alpha} + \frac{1}{2}\tilde{\psi}_{\alpha}^2\right)} \exp\left((\tilde{\mathbf{z}}_i - \tilde{\mathbf{z}}_j)^\top (\mathbf{I} + 4\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}}^{o})^{-1} (\tilde{\mathbf{z}}_i - \tilde{\mathbf{z}}_j) \right) \right]^{-1} \\
& \times 2(\mathbf{I} + 2(\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}}^{o} + \tilde{\boldsymbol{\Sigma}}_{\mathbf{w}}))^{-1} \sum_{i=1}^{N} \sum_{k=1}^{M} \left((\tilde{\mathbf{z}}_i - \tilde{\mathbf{w}}_k)(\tilde{\mathbf{z}}_i - \tilde{\mathbf{w}}_k)^\top (\mathbf{I} + 2(\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}}^{o} + \tilde{\boldsymbol{\Sigma}}_{\mathbf{w}}))^{-1} - \frac{1}{2}\mathbf{I} \right) \cdot \\
& \cdot \left[1 + \frac{\det(\mathbf{I} + 2(\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}}^{o} + \tilde{\boldsymbol{\Sigma}}_{\mathbf{w}}))^{\frac{1}{2}}}{\exp\left(\tilde{\xi}_{\beta} + \frac{1}{2}\tilde{\psi}_{\beta}^2\right)} \exp\left((\tilde{\mathbf{z}}_i - \tilde{\mathbf{w}}_k)^\top (\mathbf{I} + 2(\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}}^{o} + \tilde{\boldsymbol{\Sigma}}_{\mathbf{w}}))^{-1} (\tilde{\mathbf{z}}_i - \tilde{\mathbf{w}}_k) \right) \right]^{-1}.
\end{aligned}
$$

Therefore,

$$
\mathrm{KL} \approx \mathrm{tr}(\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}}) \left(\frac{N}{2\sigma^2} + 2\sum_{i \neq j} y_{ij} + \sum_{i=1}^{N} \sum_{k=1}^{M} x_{ik} \right) - \frac{N}{2} \log(\det(\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}})) + J(\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}}^{o})\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}} + \mathrm{Const}_{\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}}}
$$

Set $\dfrac{\partial \mathrm{KL}}{\partial \tilde{\mathbf{\Sigma}}_{\mathbf{z}}} = 0:$

$$\tilde{\mathbf{\Sigma}}_{\mathbf{z}} = \frac{N}{2}\left[\left(\frac{N}{2\sigma^2} + 2\sum_{i\neq j} y_{ij} + \sum_{i=1}^{N}\sum_{k=1}^{M} x_{ik}\right)\mathbf{I} + J(\tilde{\mathbf{\Sigma}}_{\mathbf{z}}^{o})\right]^{-1}.$$

6.4 Estimate $\tilde{\xi}_\alpha$

$$\mathrm{KL} \leq \frac{\tilde{\xi}_\alpha^2}{2\psi_\alpha^2} - \tilde{\xi}_\alpha\left(\frac{\xi_\alpha}{\psi_\alpha^2} + \sum_{i\neq j} y_{ij}\right) + \sum_{i\neq j}\log\left(1 + \exp(\tilde{\xi}_\alpha)A_{ij}\right) + \mathrm{Const}_{\tilde{\xi}_\alpha},$$

where $A_{ij} = \exp\left(\frac{1}{2}\tilde{\psi}_\alpha^2\right)\det(\mathbf{I} + 4\tilde{\mathbf{\Sigma}}_{\mathbf{z}})^{-\frac{1}{2}}\exp\left(-(\tilde{\mathbf{z}}_i - \tilde{\mathbf{z}}_j)^\top(\mathbf{I} + 4\tilde{\mathbf{\Sigma}}_{\mathbf{z}})^{-1}(\tilde{\mathbf{z}}_i - \tilde{\mathbf{z}}_j)\right)$.

Second-order Taylor series expansion of

$$f(\tilde{\xi}_\alpha^o) = \sum_{i\neq j}\log\left(1 + \exp(\tilde{\xi}_\alpha^o)A_{ij}\right),$$

evaluated at $\tilde{\xi}_\alpha = \tilde{\xi}_\alpha^o$.

We have:

$$f(\tilde{\xi}_\alpha) \approx f(\tilde{\xi}_\alpha^o) + f'(\tilde{\xi}_\alpha^o)(\tilde{\xi}_\alpha - \tilde{\xi}_\alpha^o) + \frac{1}{2}f''(\tilde{\xi}_\alpha^o)(\tilde{\xi}_\alpha - \tilde{\xi}_\alpha^o)^2,$$

where:

$$f'(\tilde{\xi}_\alpha^o) = \sum_{i\neq j}\left(1 + \exp(-\tilde{\xi}_\alpha^o)A_{ij}^{-1}\right)^{-1}$$

$$f''(\tilde{\xi}_\alpha^o) = \sum_{i\neq j}\left(1 + \exp(-\tilde{\xi}_\alpha^o)A_{ij}^{-1}\right)^{-1}\left(1 + \exp(\tilde{\xi}_\alpha^o)A_{ij}\right)^{-1}.$$

Therefore,

$$\mathrm{KL} \leq \frac{1}{2}\tilde{\xi}_\alpha^2\left(\frac{1}{\psi_\alpha^2} + f''(\tilde{\xi}_\alpha^o)\right) - \tilde{\xi}_\alpha\left(\frac{\xi_\alpha}{\psi_\alpha^2} + \sum_{i=1}^{N}\sum_{j\neq i} y_{ij} - f'(\tilde{\xi}_\alpha^o) + \tilde{\xi}_\alpha^o f''(\tilde{\xi}_\alpha^o)\right) + \mathrm{Const}_{\tilde{\xi}_\alpha}.$$

Set $\dfrac{\partial \mathrm{KL}}{\partial \tilde{\xi}_\alpha} = 0.$

$$\tilde{\xi}_\alpha = \frac{\xi_\alpha + \psi_\alpha^2(\sum_{i\neq j} y_{ij} - f'(\tilde{\xi}_\alpha^o) + \tilde{\xi}_\alpha^o f''(\tilde{\xi}_\alpha^o))}{1 + \psi_\alpha^2 f''(\tilde{\xi}_\alpha^o)}.$$

6.5 Estimate $\tilde{\psi}_\alpha^2$

$$\mathrm{KL} \leq \frac{\tilde{\psi}_\alpha^2}{2\psi_\alpha^2} - \frac{1}{2}\log(\tilde{\psi}_\alpha^2) + \sum_{i\neq j}\log\left(1 + \exp\left(\frac{1}{2}\tilde{\psi}_\alpha^2\right)A_{i,j}\right) + \mathrm{Const}_{\tilde{\psi}_\alpha^2},$$

where $A_{i,j} = \exp(\tilde{\xi}_\alpha)\det(\mathbf{I} + 4\tilde{\mathbf{\Sigma}}_{\mathbf{z}})^{-\frac{1}{2}}\exp\left(-(\tilde{\mathbf{z}}_i - \tilde{\mathbf{z}}_j)^\top(\mathbf{I} + 4\tilde{\mathbf{\Sigma}}_{\mathbf{z}})^{-1}(\tilde{\mathbf{z}}_i - \tilde{\mathbf{z}}_j)\right)$.

First-order Taylor series expansion of

$$f(\tilde{\psi}_\alpha^2) = \sum_{i \neq j} \log\left(1 + \exp\left(\frac{1}{2}\tilde{\psi}_\alpha^{2o}\right) A_{ij}\right),$$

evaluated at $\tilde{\psi}_\alpha^2 = \tilde{\psi}_\alpha^{2o}$.

We have:

$$f(\tilde{\psi}_\alpha^2) \approx f(\tilde{\psi}_\alpha^{2o}) + f'(\tilde{\psi}_\alpha^{2o})(\tilde{\psi}_\alpha^2 - \tilde{\psi}_\alpha^{2o}),$$

where

$$f'(\tilde{\psi}_\alpha^{2o}) = \sum_{i \neq j} \frac{1}{2}\left(1 + \exp\left(-\frac{1}{2}\tilde{\psi}_\alpha^{2o}\right) A_{ij}^{-1}\right)^{-1}.$$

Therefore,

$$\mathrm{KL} \approx \tilde{\psi}_\alpha^2 \left(\frac{1}{2\psi_\alpha^2} + f'(\tilde{\psi}_\alpha^{2o})\right) - \frac{1}{2}\log(\tilde{\psi}_\alpha^2) + \mathrm{Const}_{\tilde{\psi}_\alpha^2}.$$

Set $\dfrac{\partial \mathrm{KL}}{\partial \tilde{\psi}_\alpha^2} = 0:$

$$\tilde{\psi}_\alpha^2 = \left(\frac{1}{\psi_\alpha^2} + 2f'(\tilde{\psi}_\alpha^{2o})\right)^{-1}.$$

6.6 Estimate $\tilde{\xi}_\beta$

$$\mathrm{KL} \leq \frac{\tilde{\xi}_\beta^2}{2\psi_\beta^2} - \tilde{\xi}_\beta\left(\frac{\xi_\beta}{\psi_\beta^2} + \sum_{i=1}^N \sum_{k=1}^M x_{ik}\right) + \sum_{i=1}^N \sum_{k=1}^M \log\left(1 + \exp(\tilde{\xi}_\beta) A_{ik}\right) + \mathrm{Const}_{\tilde{\xi}_\beta}$$

where $A_{ik} = \exp\left(\frac{1}{2}\tilde{\psi}_\beta^2\right) \det(\mathbf{I} + 2(\tilde{\boldsymbol{\Sigma}}_\mathbf{z} + \tilde{\boldsymbol{\Sigma}}_\mathbf{w}))^{-\frac{1}{2}} \exp\left(-(\tilde{\mathbf{z}}_i - \tilde{\mathbf{w}}_k)^\top (\mathbf{I} + 2(\tilde{\boldsymbol{\Sigma}}_\mathbf{z} + \tilde{\boldsymbol{\Sigma}}_\mathbf{w}))^{-1} (\tilde{\mathbf{z}}_i - \tilde{\mathbf{w}}_k)\right)$.

Second-order Taylor series expansion of

$$f(\tilde{\xi}_\beta^o) = \sum_{i=1}^N \sum_{k=1}^M \log\left(1 + \exp(\tilde{\xi}_\beta^o) A_{ik}\right),$$

evaluated at $\tilde{\xi}_\beta = \tilde{\xi}_\beta^o$.

$$f(\tilde{\xi}_\beta) \approx f(\tilde{\xi}_\beta^o) + f'(\tilde{\xi}_\beta^o)(\tilde{\xi}_\beta - \tilde{\xi}_\beta^o) + \frac{1}{2} f''(\tilde{\xi}_\beta^o)(\tilde{\xi}_\beta - \tilde{\xi}_\beta^o)^2$$

where:

$$f'(\tilde{\xi}_\beta^o) = \sum_{i=1}^N \sum_{k=1}^M \left(1 + \exp(-\tilde{\xi}_\beta^o) A_{ik}^{-1}\right)^{-1},$$

$$f''(\tilde{\xi}_\beta^o) = \sum_{i=1}^N \sum_{k=1}^M \left(1 + \exp(-\tilde{\xi}_\beta^o) A_{ik}^{-1}\right)^{-1} \left(1 + \exp(\tilde{\xi}_\beta^o) A_{ik}\right)^{-1}.$$

Therefore,

$$\mathrm{KL} \le \frac{1}{2}\tilde{\xi}_\beta^2 \left(\frac{1}{\psi_\beta^2} + f''(\tilde{\xi}_\beta^o)\right) - \tilde{\xi}\left(\frac{\xi_\beta}{\psi_\beta^2} + \sum_{i=1}^N \sum_{k=1}^M x_{ik} - f'(\tilde{\xi}_\beta^o) + \tilde{\xi}_\beta^o f''(\tilde{\xi}_\beta^o)\right) + \mathrm{Const}_{\tilde{\xi}_\beta}.$$

Set $\dfrac{\partial \mathrm{KL}}{\partial \tilde{\xi}_\beta} = 0$:

$$\tilde{\xi}_\beta = \frac{\xi_\beta + \psi_\beta^2(\sum_{i=1}^N \sum_{k=1}^M x_{ik} - f'(\tilde{\xi}_\beta^o) + \tilde{\xi}_\beta^o f''(\tilde{\xi}_\beta^o))}{1 + \psi_\beta^2 f''(\tilde{\xi}_\beta^o)}.$$

6.7 Estimate $\tilde{\psi}_\beta^2$

$$\mathrm{KL} \le \frac{\tilde{\psi}_\beta^2}{2\psi_\beta^2} - \frac{1}{2}\log(\tilde{\psi}_\beta^2) + \sum_{i=1}^N \sum_{k=1}^M \log\left(1 + \exp\left(\frac{1}{2}\tilde{\psi}_\beta^2\right) A_{ik}\right) + \mathrm{Const}_{\tilde{\psi}_\beta^2},$$

where $A_{ik} = \exp(\tilde{\xi}_\beta)\det(\mathbf{I}+2(\tilde{\mathbf{\Sigma}}_\mathbf{z}+\tilde{\mathbf{\Sigma}}_\mathbf{w}))^{-\frac{1}{2}} \exp\Big(-(\tilde{\mathbf{z}}_i - \tilde{\mathbf{w}}_k)^\top (\mathbf{I} + 2(\tilde{\mathbf{\Sigma}}_\mathbf{z} + \tilde{\mathbf{\Sigma}}_\mathbf{w}))^{-1}(\tilde{\mathbf{z}}_i - \tilde{\mathbf{w}}_k)\Big)$.

First-order Taylor series expansion of:

$$f(\tilde{\psi}_\beta^2) = \sum_{i=1}^N \sum_{k=1}^M \log\left(1 + \exp\left(\frac{1}{2}\tilde{\psi}_\beta^{2o}\right) A_{ik}\right),$$

evaluated at $\tilde{\psi}_\beta^2 = \tilde{\psi}_\beta^{2o}$.

We have:

$$f(\tilde{\psi}_\beta^2) \approx f(\tilde{\psi}_\beta^{2o}) + f'(\tilde{\psi}_\beta^{2o})(\tilde{\psi}_\beta^2 - \tilde{\psi}_\beta^{2o}).$$

where

$$f'(\tilde{\psi}_\beta^{2o}) = \sum_{i=1}^N \sum_{k=1}^M \frac{1}{2}\left(1 + \exp\left(-\frac{1}{2}\tilde{\psi}_\beta^{2o}\right) A_{ik}^{-1}\right)^{-1}.$$

Therefore,

$$\mathrm{KL} \approx \tilde{\psi}_\beta^2 \left(\frac{1}{2\psi_\beta^2} + f'(\tilde{\psi}_\beta^{2o})\right) - \frac{1}{2}\log(\tilde{\psi}_\beta^2) + \mathrm{Const}_{\tilde{\psi}_\beta^2}.$$

Set $\dfrac{\partial \mathrm{KL}}{\partial \tilde{\psi}_\beta^2} = 0.$

$$\tilde{\psi}_\beta^2 = \left(\frac{1}{\psi_\beta^2} + 2f'(\tilde{\psi}_\beta^{2o})\right)^{-1}.$$

6.8 Estimate $\tilde{\mathbf{w}}_k$

$$\mathrm{KL} \leq \mathrm{Const}_{\tilde{\mathbf{w}}_k} + \tilde{\mathbf{w}}_k^\top \tilde{\mathbf{w}}_k \left(\frac{1}{2\sigma^2} + \sum_{i=1}^{N} x_{ik} \right) - 2\tilde{\mathbf{w}}_k^\top \left(\sum_{i=1}^{N} x_{ik} \tilde{\mathbf{z}}_i \right)$$
$$+ \sum_{k=1}^{M} \log \left(1 + \frac{\exp\left(\tilde{\xi}_\beta + \frac{1}{2}\tilde{\psi}_\beta^2\right)}{\det(\mathbf{I} + 2(\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}} + \tilde{\boldsymbol{\Sigma}}_{\mathbf{w}}))^{\frac{1}{2}}} \exp\Big(-(\tilde{\mathbf{z}}_i - \tilde{\mathbf{w}}_k)^\top (\mathbf{I} + 2(\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}} + \tilde{\boldsymbol{\Sigma}}_{\mathbf{w}}))^{-1} (\tilde{\mathbf{z}}_i - \tilde{\mathbf{w}}_k) \Big) \right).$$

Second-order Taylor series expansion approximation of

$$f(\tilde{\mathbf{z}}_i) = \frac{1}{2} \sum_{k=1}^{M} \log \left(1 + \frac{\exp\left(\tilde{\xi}_\beta + \frac{1}{2}\tilde{\psi}_\beta^2\right)}{\det(\mathbf{I} + 2(\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}} + \tilde{\boldsymbol{\Sigma}}_{\mathbf{w}}))^{\frac{1}{2}}} \exp\Big(-(\tilde{\mathbf{z}}_i - \tilde{\mathbf{w}}_k)^\top (\mathbf{I} + 2(\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}} + \tilde{\boldsymbol{\Sigma}}_{\mathbf{w}}))^{-1} (\tilde{\mathbf{z}}_i - \tilde{\mathbf{w}}_k) \Big) \right).$$

Therefore,

$$f(\tilde{\mathbf{w}}_k) \approx f(\tilde{\mathbf{w}}_k^o) + (\tilde{\mathbf{w}}_k - \tilde{\mathbf{w}}_k^o)^\top G(\tilde{\mathbf{w}}_k^o) + \frac{1}{2} (\tilde{\mathbf{w}}_k - \tilde{\mathbf{w}}_k^o)^\top H(\tilde{\mathbf{w}}_k^o) (\tilde{\mathbf{w}}_k - \tilde{\mathbf{w}}_k^o).$$

Let's find the gradient G and the Hessian matrix H of f evaluated at $\tilde{\mathbf{w}}_k = \tilde{\mathbf{w}}_k^o$.

$$G(\tilde{\mathbf{w}}_k^o) = -\left(\mathbf{I} + 2(\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}} + \tilde{\boldsymbol{\Sigma}}_{\mathbf{w}})\right)^{-1} \sum_{k=1}^{M} (\tilde{\mathbf{z}}_i - \tilde{\mathbf{w}}_k^o)$$
$$\times \left[1 + \frac{\det(\mathbf{I} + 2(\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}} + \tilde{\boldsymbol{\Sigma}}_{\mathbf{w}}))^{\frac{1}{2}}}{\exp\left(\tilde{\xi}_\beta + \frac{1}{2}\tilde{\psi}_\beta^2\right)} \exp\Big((\tilde{\mathbf{z}}_i - \tilde{\mathbf{w}}_k^o)^\top (\mathbf{I} + 2(\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}} + \tilde{\boldsymbol{\Sigma}}_{\mathbf{w}}))^{-1} (\tilde{\mathbf{z}}_i - \tilde{\mathbf{w}}_k^o) \Big) \right]^{-1}.$$

$$H(\tilde{\mathbf{w}}_k^o) = -\left(\mathbf{I} + 2(\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}} + \tilde{\boldsymbol{\Sigma}}_{\mathbf{w}})\right)^{-1} \sum_{k=1}^{M} \left[1 + \frac{\det(\mathbf{I} + 2(\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}} + \tilde{\boldsymbol{\Sigma}}_{\mathbf{w}}))^{\frac{1}{2}}}{\exp\left(\tilde{\xi}_\beta + \frac{1}{2}\tilde{\psi}_\beta^2\right)} \exp\Big((\tilde{\mathbf{z}}_i - \tilde{\mathbf{w}}_k^o)^\top (\mathbf{I} + 2(\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}} + \tilde{\boldsymbol{\Sigma}}_{\mathbf{w}}))^{-1} (\tilde{\mathbf{z}}_i - \tilde{\mathbf{w}}_k^o) \Big) \right]^{-1}$$
$$\times \left[\mathbf{I} - \frac{2(\tilde{\mathbf{z}}_i - \tilde{\mathbf{w}}_k^o)(\tilde{\mathbf{z}}_i - \tilde{\mathbf{w}}_k^o)^\top (\mathbf{I} + 2(\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}} + \tilde{\boldsymbol{\Sigma}}_{\mathbf{w}}))^{-1}}{1 + \frac{\exp\left(\tilde{\xi}_\beta + \frac{1}{2}\tilde{\psi}_\beta^2\right)}{\det(\mathbf{I} + 2(\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}} + \tilde{\boldsymbol{\Sigma}}_{\mathbf{w}}))^{\frac{1}{2}}} \exp\Big(-(\tilde{\mathbf{z}}_i - \tilde{\mathbf{w}}_k^o)^\top (\mathbf{I} + 2(\tilde{\boldsymbol{\Sigma}}_{\mathbf{z}} + \tilde{\boldsymbol{\Sigma}}_{\mathbf{w}}))^{-1} (\tilde{\mathbf{z}}_i - \tilde{\mathbf{w}}_k^o) \Big)} \right].$$

Therefore,

$$\mathrm{KL} \approx \tilde{\mathbf{w}}_k^\top \left[\left(\frac{1}{2\sigma^2} + \sum_{i=1}^{N} x_{ik} \right) \mathbf{I} + H(\tilde{\mathbf{w}}_k^o) \right] \tilde{\mathbf{w}}_k$$
$$- 2\tilde{\mathbf{w}}_k^\top \left[\sum_{i=1}^{N} x_{ik} \tilde{\mathbf{z}}_i - G(\tilde{\mathbf{w}}_k^o) + H(\tilde{\mathbf{w}}_k^o) \tilde{\mathbf{w}}_k^o \right] + \mathrm{Const}_{\tilde{\mathbf{w}}_k}.$$

Set $\dfrac{\partial \mathrm{KL}}{\partial \tilde{\mathbf{w}}_k} = 0.$

$$\tilde{\mathbf{w}}_k = \left[\left(\frac{1}{2\sigma^2} + \sum_{i=1}^{N} x_{ik} \right) \mathbf{I} + H(\tilde{\mathbf{w}}_k^o) \right]^{-1} \left[\sum_{i=1}^{N} x_{ik} \tilde{\mathbf{z}}_i - G(\tilde{\mathbf{w}}_k^o) + H(\tilde{\mathbf{w}}_k^o) \tilde{\mathbf{w}}_k^o \right].$$

6.9 Estimate $\tilde{\mathbf{\Sigma}}_\mathbf{w}$

$$\begin{aligned} \mathrm{KL} \le \; & \mathrm{Const}_{\tilde{\mathbf{\Sigma}}_\mathbf{w}} + \mathrm{tr}(\tilde{\mathbf{\Sigma}}_\mathbf{w}) \left(\frac{M}{2\sigma^2} + \sum_{i=1}^{N} \sum_{k=1}^{M} x_{ik} \right) - \frac{M}{2} \log(\det(\tilde{\mathbf{\Sigma}}_\mathbf{w})) \\ & + \sum_{i=1}^{N} \sum_{k=1}^{M} \log \left(1 + \frac{\exp\left(\tilde{\xi}_\beta + \frac{1}{2} \tilde{\psi}_\beta^2 \right)}{\det(\mathbf{I} + 2(\tilde{\mathbf{\Sigma}}_\mathbf{z} + \tilde{\mathbf{\Sigma}}_\mathbf{w}))^{\frac{1}{2}}} \exp\left(-(\tilde{\mathbf{z}}_i - \tilde{\mathbf{w}}_k)^\top (\mathbf{I} + 2(\tilde{\mathbf{\Sigma}}_\mathbf{z} + \tilde{\mathbf{\Sigma}}_\mathbf{w}))^{-1} (\tilde{\mathbf{z}}_i - \tilde{\mathbf{w}}_k) \right) \right). \end{aligned}$$

First-order Taylor series expansion approximation of:

$$f(\tilde{\mathbf{\Sigma}}_\mathbf{w}) = \sum_{i=1}^{N} \sum_{k=1}^{M} \log \left(1 + \frac{\exp\left(\tilde{\xi}_\beta + \frac{1}{2} \tilde{\psi}_\beta^2 \right)}{\det(\mathbf{I} + 2(\tilde{\mathbf{\Sigma}}_\mathbf{z} + \tilde{\mathbf{\Sigma}}_\mathbf{w}))^{\frac{1}{2}}} \exp\left(-(\tilde{\mathbf{z}}_i - \tilde{\mathbf{w}}_k)^\top (\mathbf{I} + 2(\tilde{\mathbf{\Sigma}}_\mathbf{z} + \tilde{\mathbf{\Sigma}}_\mathbf{w}))^{-1} (\tilde{\mathbf{z}}_i - \tilde{\mathbf{w}}_k) \right) \right)$$

$$f(\tilde{\mathbf{\Sigma}}_\mathbf{w}) \approx f(\tilde{\mathbf{\Sigma}}_\mathbf{w}^o) + J(\tilde{\mathbf{\Sigma}}_\mathbf{w}^o)(\tilde{\mathbf{\Sigma}}_\mathbf{w} - \tilde{\mathbf{\Sigma}}_\mathbf{w}^o).$$

where J is the Jacobian matrix of f evaluated at $\tilde{\mathbf{\Sigma}}_\mathbf{w} = \tilde{\mathbf{\Sigma}}_\mathbf{w}^o$.

$$\begin{aligned} J(\tilde{\mathbf{\Sigma}}_\mathbf{w}^o) = \; & 2(\mathbf{I} + 2(\tilde{\mathbf{\Sigma}}_\mathbf{z} + \tilde{\mathbf{\Sigma}}_\mathbf{w}^o))^{-1} \sum_{i=1}^{N} \sum_{k=1}^{M} \left((\tilde{\mathbf{z}}_i - \tilde{\mathbf{w}}_k)(\tilde{\mathbf{z}}_i - \tilde{\mathbf{w}}_k)^\top (\mathbf{I} + 2(\tilde{\mathbf{\Sigma}}_\mathbf{z} + \tilde{\mathbf{\Sigma}}_\mathbf{w}^o))^{-1} - \frac{1}{2}\mathbf{I} \right) \cdot \\ & \cdot \left[1 + \frac{\det(\mathbf{I} + 2(\tilde{\mathbf{\Sigma}}_\mathbf{z} + \tilde{\mathbf{\Sigma}}_\mathbf{w}^o))^{\frac{1}{2}}}{\exp\left(\tilde{\xi}_\beta + \frac{1}{2} \tilde{\psi}_\beta^2 \right)} \exp\left((\tilde{\mathbf{z}}_i - \tilde{\mathbf{w}}_k)^\top (\mathbf{I} + 2(\tilde{\mathbf{\Sigma}}_\mathbf{z} + \tilde{\mathbf{\Sigma}}_\mathbf{w}^o))^{-1} (\tilde{\mathbf{z}}_i - \tilde{\mathbf{w}}_k) \right) \right]^{-1}. \end{aligned}$$

Therefore,

$$\mathrm{KL} \approx \mathrm{tr}(\tilde{\mathbf{\Sigma}}_\mathbf{w}) \left(\frac{M}{2\sigma^2} + \sum_{i=1}^{N} \sum_{k=1}^{M} x_{ik} \right) - \frac{M}{2} \log(\det(\tilde{\mathbf{\Sigma}}_\mathbf{w})) + J(\tilde{\mathbf{\Sigma}}_\mathbf{w}^o)\tilde{\mathbf{\Sigma}}_\mathbf{w} + \mathrm{Const}_{\tilde{\mathbf{\Sigma}}_\mathbf{w}}.$$

Set $\dfrac{\partial \mathrm{KL}}{\partial \tilde{\mathbf{\Sigma}}_\mathbf{w}} = 0:$

$$\tilde{\mathbf{\Sigma}}_\mathbf{w} = \frac{M}{2} \left[\left(\frac{M}{2\sigma^2} + \sum_{i=1}^{N} \sum_{k=1}^{M} x_{ik} \right) \mathbf{I} + J(\tilde{\mathbf{\Sigma}}_\mathbf{w}^o) \right]^{-1}.$$