Bayesian Robust Attributed Graph Clustering:

Joint Learning of Partial Anomalies and Group Structure

Supplementary Material

1 Experiments

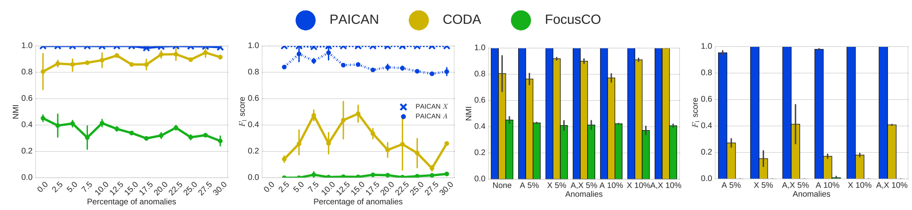
1.1 Experimental setup details

For all methods we provide the true number K of clusters to detect. CODA is a non-deterministic algorithm that is highly sensitive to initialization and parameter choice. Following [2], we tried the values of $\{0.05, 0.1, 0.5\}$ for λ, perform several restarts for each of them and report only the highest NMI achieved. Moreover, CODA requires the percentage of anomalous nodes, for which we provide the true values. FocusCO is a semi-supervised approach; we have to provide example nodes that belong to the same single cluster. We pick the best possible scenario, i.e. we run FocusCO K number of times and we provide *all* nodes from a given cluster. We then pick the run which gave us the highest NMI. SIAN and LSBM are also executed multiple times and we pick the solution achieving highest NMI. For our approach, we simply perform several restarts with random initialization and pick the one that gives us the highest *likelihood*. PICS and BAGC are deterministic.

1.2 Blocky clusters: Robustness and anomaly detection

In Section 5.1 of the paper we mention that if we perform similar analysis for the unrealistic case of 'blocky' clusters (i.e. the degree distribution does not show a power-law) FocusCO and CODA perform relatively better.

In Fig. 1(a) we can see that CODA performs relatively well achieving high NMI score for the blocky clusters even though it can not detect all the anomalies as shown in Fig. 1(b). FocusCO although performs better compared to the power-law degree distributed graphs still has poor performance in comparison. PAICAN consistently outperforms both methods. We can draw similar conculusions for the case when we are generating either only graph anomalies (A), attribute anomalies (X), or complete anomalies (A, X) shown in Figs. 1(c) and 1(d).



(a) NMI (b) F_1 score (c) NMI (d) F_1 score

Figure 1: Clustering and anomaly detection performance when increasing the percentage of anomalies on synthetic data with *blocky clusters*. PAICAN is very robust obtaining highest results.

1.3 Case study: Connectivity patterns

In this experiment we run our method on the SOCIALPAPERS and inspect the inferred block connectivity patterns $\boldsymbol{\eta}$. The nature of this dataset yields non-trivial block structure where some clusters have significant number of edges between each other. In other words we have relatively high values at multiple places in $\boldsymbol{\eta}$ rather than just on the diagonal.

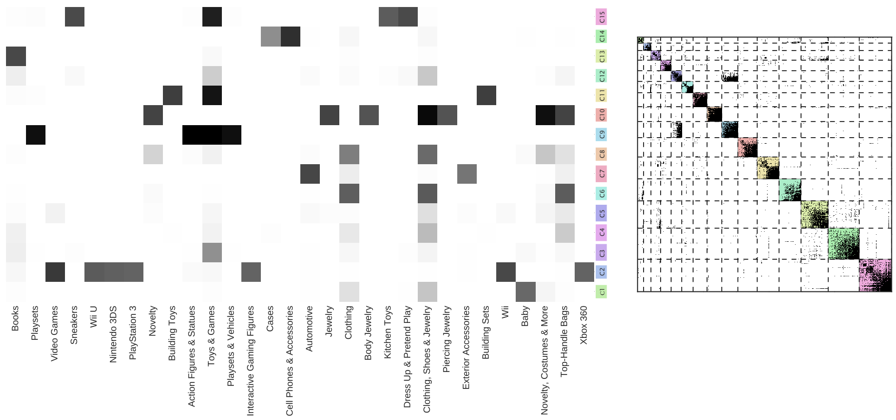
For example our method detects a cluster where almost all of the papers are from the subject 'neurology'. We observe that this cluster most likely connects to two other clusters with main subjects 'diagnostic imaging' and 'psychiatry' respectively. These connection patterns are coherent and indicative of users who tend to mention papers from the broader area of neuropsychiatry. Similarly we discover a cluster of papers about 'audiology' with preferred connections to a cluster about 'speech language pathology'.

1.4 Case study: Clustering

To enable visual inspection of the clustering, we select a small subset ($N = 1549, E = 36934, D = 661$) of the AMAZON dataset. The results for $K = 15$ are visualized in Fig. 2 and Fig. 3(a) . The learned topic distribution $\mathbf{t}$ is shown in Fig. 2(a); for visualization we only plot dimensions where $t_{dk} > 0.5$ for at least one cluster. Intuitively, this plot shows the 'active' categories for each cluster. For example the products in cluster $C2$ have the following most active categories [Wii U, Nintendo 3DS, PlayStation 3, Xbox 360] clearly showing a coherent cluster of products related to gaming consoles. Similarly, inspecting the topics of $C10$ shows products about jewelery and $C14$ cell phone cases related products.

The adjacency matrix in Fig. 2(b) reveals that the network has mostly typical clusters – with most edges within the clusters and few edges between the clusters. Notable is one off-diagonal entry between clusters $C9$ and $C5$. Interestingly, this entry describes co-purchase behavior between the cluster with topic 'playsets, toys, actionfigures' and the cluster 'clothing, bags'. Hence, probably indicating co-buying behavior of families where products for their kids are bought together with other products. PAICAN can easily detect such kind of network topology.

Finally, the graph embedding in Fig. 3(a) visually confirms the good clustering



(a) Topics of the clusters

(b) Adjacency matrix

Figure 2: Clustering in the Amazon dataset: cluster topics and adjacency matrix. Colors indicate clusters.

structure and shows that PAICAN can easily handle non-trivial degree distributions. We also observe partial graph anomalies – marked bigger in size – connecting to several unrelated (according to $\boldsymbol{\eta}$) clusters.

1.5 Case study: Partial anomalies

For this case study we run our method on the DBLP dataset. Since the data is too large to visualize and we have no anomaly ground truth to evaluate the validity of our results we pick some of the detected partial anomalous nodes and inspect their ego network and attributes.

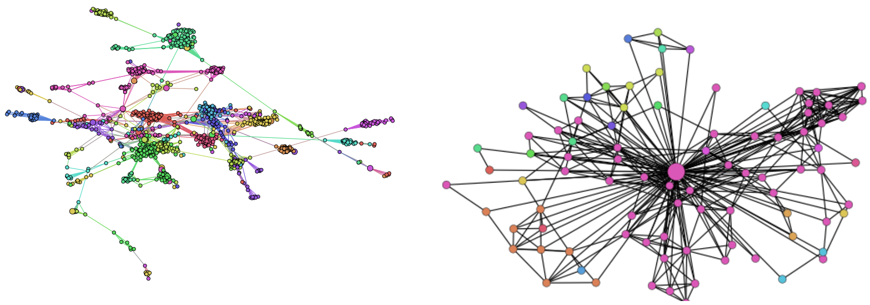
In Fig. 3(b) we show the ego network for a node that has been marked partially anomalous in attribute space corresponding to Srinivasan Parthasarathy. As we can see from the ego-network he fits nicely in graph space since most of his neighbors belong to the same cluster. However, as we discussed in the main paper in, he is an obvious anomaly w.r.t. attribute space. Overall, all these case studies indicate that PAICAN is able to extract interesting knowledge from attributed graphs.

2 Proofs and derivations

2.1 Terms of the ELBO

Given our model, the ELBO decomposes as follows:

$$\underbrace{\mathbb{E}_q[\log p(A|\mathbf{z},\mathbf{c},\boldsymbol{\eta},\eta_{bg},\eta_{bb},\boldsymbol{\theta},\widetilde{\boldsymbol{\theta}})]}_{:=\mathcal{L}_{\mathcal{A}}} + \underbrace{\mathbb{E}_q[\log p(X|\mathbf{z},\mathbf{c},\mathbf{t})]}_{:=\mathcal{L}_{\mathcal{X}}} + \mathbb{E}_q[\log p(\mathbf{z}|\mathbf{c},\boldsymbol{\pi})] + \mathbb{E}_q[\log p(\mathbf{c}|\boldsymbol{\rho})] - \mathbb{E}_q[\log q(\mathbf{z},\mathbf{c})] \quad (1)$$



(a) Clustering in the Amazon dataset. Colors indicate clusters. Bigger nodes indicate partial graph anomalies. Best viewed on screen.

(b) Ego network of a partial anomaly in attribute space for DBLP data. Colors indicate clusters.

Figure 3: Graph embeddings for Amazon and DBLP data.

The last four terms are straightforward and can all be evaluated in linear time w.r.t. the number of nodes and dimensions.

$$\mathbb{E}_q[\log p(\mathbf{c}|\boldsymbol{\rho})] = \sum_i \sum_{m=0}^{3} \phi_{im} \log(\rho_m)$$

$$\mathbb{E}_q[\log p(\mathbf{z}|\mathbf{c}, \boldsymbol{\pi})] = \sum_i \sum_k \psi_{ik}(1 - \phi_{i3}) log(\pi_k)$$

$$\mathbb{E}_q[\log q(\mathbf{z}, \mathbf{c})] = \sum_i \sum_k \psi_{ik} \log(\psi_{ik}) + \sum_i \sum_{m=0}^{3} \phi_{im} \log(\phi_{im})$$

$$\mathcal{L}_{\mathcal{X}} = \sum_i \sum_k \psi_{ik} \phi_{i0}^X \left(\sum_d X_{id} \log(t_{dk}) + (1 - X_{id}) \log(1 - t_{dk}) \right) + \sum_i \phi_{i1}^X D \log(0.5)$$

2.2 Proof of Eq. (4)

We will show that $\sum_i \phi_{i0}^A \theta_i \sum_l \psi_{il} \eta_{kl} = 1, \forall k$. We are using this identity only in the E-step of our variational EM. That means we can substitute our MLE solution for the parameters that we got in the M-step.

Let's start by plugging in the solution for η_{kl} and substituting D_k^G, we get:

$$\sum_i \phi_{i0}^A \theta_i \sum_l \psi_{il} \eta_{kl} = \sum_l \psi_{il} \sum_i \phi_{i0}^A \theta_i \frac{m_{kl}}{D_k^G D_l^G} =$$

$$\sum_l \sum_i \phi_{i0}^A \theta_i \psi_{il} \frac{m_{kl}}{\left(\sum_i \phi_{i0}^A \theta_i \psi_{ik}\right)\left(\sum_i \phi_{i0}^A \theta_i \psi_{il}\right)} = \frac{\sum_l m_{kl}}{\sum_i \phi_{i0}^A \theta_i \psi_{ik}}$$

Let's substitute now θ_i and m_{kl} and take advantage of $\sum_k \psi_k C = C$ for any constant C that does not depend on k:

$$\frac{\sum_l m_{kl}}{\sum_i \phi_{i0}^A \theta_i \psi_{ik}} = \frac{\sum_l \sum_{i \neq j} A_{ij} \phi_{i0}^A \phi_{j0}^A \psi_{ik} \psi_{jl}}{\sum_i \phi_{i0}^A (\sum_{j \neq i} A_{ij} \phi_{j0}^A) \psi_{ik}} = \frac{\sum_l \psi_{jl} \sum_{i \neq j} A_{ij} \phi_{i0}^A \phi_{j0}^A \psi_{ik}}{\sum_{i \neq j} A_{ij} \phi_{i0}^A \phi_{j0}^A \psi_{ik}} = \frac{\sum_{i \neq j} A_{ij} \phi_{i0}^A \phi_{j0}^A \psi_{ik}}{\sum_{ij} A_{ij} \phi_{i0}^A \phi_{j0}^A \psi_{ik}} = 1$$

2.3 Calculation of the E-Step

Following [1] (Ch. 10) the optimal variational distribution is

$$q^*(z_i) \propto \exp(\mathbb{E}_{q\setminus z_i}[\log p(A, X, \mathbf{z}, \mathbf{c}| \dots)])$$

Thus to derive the optimal variational parameters ψ_{ik} for the cluster assignments, we have to keep only the terms of the ELBO that include ψ_{ik} and disregard the constants. After rearranging we obtain (for the general case of graphs including potential self-loops):

$$\psi_{ik}^{new} \propto \exp\Bigg(\phi_{i0}^A\Bigg[\sum_{j\in\mathcal{N}_i}\phi_{j0}^A\sum_l \psi_{jl}\log(\theta_i\theta_j\eta_{kl}) - \theta_i(1-\theta_i\phi_{i0}^A\sum_l\psi_{il}\eta_{kl}) - \frac{1}{2}\theta_i^2\eta_{kk}$$
$$+ A_{ii}\log(\frac{1}{2}\theta_i^2\eta_{kk})\Bigg] + \phi_{i0}^X\Bigg[\sum_d X_{id}\log(t_{dk}) + (1-X_{id})\log(1-t_{dk})\Bigg] + (1-\phi_{i3})\log(\pi_k)\Bigg) \tag{2}$$

In the case when the graph contains no self-loops ($A_{ii} = 0$) we obtain Eq. (9) in the paper.

Similarly, for the anomaly assignments ϕ_{im}, when including the self loops, the variables $\hat{\phi}_{i0}^A$ and $\hat{\phi}_{i1}^A$ become:

$$\begin{aligned}
\hat{\phi}_{i0}^A &= \sum_{j\in\mathcal{N}_i}\phi_{j0}^A\sum_{kl}\psi_{ik}\psi_{jl}\log(\theta_i\theta_j\eta_{kl}) - \theta_i(1-\theta_i\phi_{i0}^A\sum_{kl}\psi_{ik}\psi_{il}\eta_{kl})\\
&+ \sum_{j\in\mathcal{N}_i}\phi_{j1}^A\log(\widetilde{\theta}_j\eta_{bg}) - \eta_{bg}(\widetilde{\theta}^B - \phi_{i1}^A\widetilde{\theta}_i) - \frac{1}{2}\theta_i^2\sum_k\psi_{ik}\eta_{kk} + \sum_k\psi_{ik}A_{ii}\log(\frac{1}{2}\theta_i^2\eta_{kk})\\
\hat{\phi}_{i1}^A &= \log(\widetilde{\theta}_i\eta_{bg})\sum_{j\in\mathcal{N}_i}\phi_{j0}^A - \eta_{bg}\widetilde{\theta}_i(g-\phi_{i0}^A) + \sum_{j\in\mathcal{N}_i}\phi_{j1}^A\log(\widetilde{\theta}_i\widetilde{\theta}_j\eta_{bb})\\
&- \widetilde{\theta}_i\eta_{bb}(\widetilde{\theta}^B - \phi_{i1}^A\widetilde{\theta}_i) - \frac{1}{2}\widetilde{\theta}_i^2\eta_{bb} + A_{ii}\log(\frac{1}{2}\widetilde{\theta}_i^2\eta_{bb})
\end{aligned}$$

2.4 Reformulation of the ELBO for the M-Step

We will simplify $\mathcal{L}_\mathcal{A}$ – the ELBO term with regards to graph space. We start with the definition:

$$\begin{aligned}
\mathcal{L}_\mathcal{A} &= \sum_{i<j}\sum_{k,l}\psi_{ik}\psi_{jl}\phi_{i0}^A\phi_{j0}^A A_{ij}\log(\theta_i\theta_j\eta_{kl}) - \psi_{ik}\psi_{jl}\phi_{i0}^A\phi_{j0}^A\theta_i\theta_j\eta_{kl}\\
&+ \sum_{i<j}\phi_{i1}^A\phi_{j0}^A A_{ij}\log(\widetilde{\theta}_i\eta_{bg}) - \phi_{i1}^A\phi_{j0}^A\widetilde{\theta}_i\eta_{bg}\\
&+ \sum_{i<j}\phi_{i0}^A\phi_{j1}^A A_{ij}\log(\widetilde{\theta}_j\eta_{bg}) - \phi_{i0}^A\phi_{j1}^A\widetilde{\theta}_j\eta_{bg} + \sum_{i<j}\phi_{i1}^A\phi_{j1}^A A_{ij}\log(\widetilde{\theta}_i\widetilde{\theta}_j\eta_{bb}) - \phi_{i1}^A\phi_{j1}^A\widetilde{\theta}_i\widetilde{\theta}_j\eta_{bb}\\
&+ \sum_i\sum_k\psi_{ik}\phi_{i0}^A A_{ii}\log(\frac{1}{2}\theta_i^2\eta_{kk}) - \psi_{ik}\phi_{i0}^A\frac{1}{2}\theta_i^2\eta_{kk} + \sum_i\phi_{i1}^A A_{ii}\log(\frac{1}{2}\widetilde{\theta}_i^2\eta_{bb}) - \phi_{i1}^A\frac{1}{2}\widetilde{\theta}_i^2\eta_{bb}
\end{aligned} \tag{3}$$

Now we will consider the terms involving $\boldsymbol{\theta}$ and $\widetilde{\boldsymbol{\theta}}$. Looking at terms involving $\boldsymbol{\theta}$ in (3) a, and taking advantage of symmetry, we can rewrite them as:

$$\begin{aligned}\mathcal{L}_{\boldsymbol{\theta}} &= \sum_{i\neq j}\sum_{k,l}\psi_{ik}\psi_{jl}\phi^A_{i0}\phi^A_{j0}A_{ij}\log(\theta_i) + \frac{1}{2}\sum_{i\neq j}\sum_{k,l}\psi_{ik}\psi_{jl}\phi^A_{i0}\phi^A_{j0}A_{ij}\log(\eta_{kl}) \\ &- \frac{1}{2}\sum_{i\neq j}\sum_{k,l}\psi_{ik}\psi_{jl}\phi^A_{i0}\phi^A_{j0}\theta_i\theta_j\eta_{kl} + \sum_i\sum_k\psi_{ik}\phi^A_{i0}A_{ii}\log(\frac{1}{2}\theta_i^2\eta_{kk}) - \psi_{ik}\phi^A_{i0}\frac{1}{2}\theta_i^2\eta_{kk}\end{aligned}$$

Which we can rewrite using the definitions of m_{kl} and D^G_k:

$$\begin{aligned}\mathcal{L}_{\boldsymbol{\theta}} &= \sum_i\log(\theta_i)\phi^A_{i0}d^G_i + \frac{1}{2}\sum_{k,l}m_{kl}log(\eta_{kl}) - \frac{1}{2}D^G_kD^G_l\eta_{kl} \\ &+ \frac{1}{2}\sum_i\sum_{k,l}\psi_{ik}\psi_{il}\theta_i^2\phi^A_{i0}(\phi^A_{i0}\eta_{kl} - \eta_{kk}) + \sum_i\phi^A_{i0}A_{ii}\log(\frac{1}{2}\theta_i^2)\end{aligned}$$

Now let's look at the terms involving $\widetilde{\boldsymbol{\theta}}$. Again taking advantage of symmetry we can rewrite it as:

$$\begin{aligned}\mathcal{L}_{\widetilde{\boldsymbol{\theta}}} &= \sum_{i\neq j}\phi^A_{i1}\phi^A_{j0}A_{ij}\log(\widetilde{\theta}_i) + \phi^A_{i1}\phi^A_{j0}A_{ij}\log(\eta_{bg}) - \phi^A_{i1}\phi^A_{j0}\widetilde{\theta}_i\eta_{bg} \\ &+ \sum_{i\neq j}\phi^A_{i1}\phi^A_{j1}A_{ij}\log(\widetilde{\theta}_i) + \frac{1}{2}\phi^A_{i1}\phi^A_{j1}A_{ij}\log(\eta_{bb}) - \frac{1}{2}\phi^A_{i1}\phi^A_{j1}\widetilde{\theta}_i\widetilde{\theta}_j\eta_{bb} \\ &+ \sum_i\phi^A_{i1}A_{ii}\log(\frac{1}{2}\widetilde{\theta}_i^2\eta_{bb}) - \phi^A_{i1}\frac{1}{2}\widetilde{\theta}_i^2\eta_{bb}\end{aligned}$$

Which we can then rewrite using the definitions of m_{bg}, m_{bb} as:

$$\begin{aligned}\mathcal{L}_{\widetilde{\boldsymbol{\theta}}} &= \sum_i\sum_{j\in\mathcal{N}_i}\phi^A_{i1}\phi^A_{j0}\log(\widetilde{\theta}_i) + m_{bg}\log(\eta_{bg}) - \sum_{i\neq j}\phi^A_{i1}\phi^A_{j0}\widetilde{\theta}_i\eta_{bg}\sum_i\sum_{j\in\mathcal{N}_i}\phi^A_{i1}\phi^A_{j1}\log(\widetilde{\theta}_i) \\ &\frac{1}{2}m_{bb}\log(\eta_{bb}) - \frac{1}{2}\sum_{i\neq j}\phi^A_{i1}\phi^A_{j1}\widetilde{\theta}_i\widetilde{\theta}_j\eta_{bb} + \sum_i\phi^A_{i1}A_{ii}\log(\frac{1}{2}\widetilde{\theta}_i^2\eta_{bb}) - \phi^A_{i1}\frac{1}{2}\widetilde{\theta}_i^2\eta_{bb}\end{aligned}$$

Plugging in D^B and further simplifying we arrive at:

$$\begin{aligned}\mathcal{L}_{\widetilde{\boldsymbol{\theta}}} &= \sum_i\phi^A_{i1}d_i\log(\widetilde{\theta}_i) + m_{bg}\log(\eta_{bg}) + \frac{1}{2}m_{bb}\log(\eta_{bb}) - \frac{1}{2}D^BD^B\eta_{bb} \\ &- D^B\sum_i(\sum_j\phi^A_{j0} - \phi^A_{i0})\eta_{bg} + \frac{1}{2}\sum_i\widetilde{\theta}_i^2\eta_{bb}\phi^A_{i1}(\phi^A_{i1} - 1) + \sum_i\phi^A_{i1}A_{ii}\log(\frac{1}{2}\widetilde{\theta}_i^2)\end{aligned}$$

Joining the terms regarding $\boldsymbol{\theta}$ and $\widetilde{\boldsymbol{\theta}}$, plugging in the definition of g and rearranging we the complete log-likelihood w.r.t. to graph space:

$$\begin{aligned}\mathcal{L}_{\mathcal{A}} &= \frac{1}{2}\sum_{k,l}m_{kl}log(\eta_{kl}) - \frac{1}{2}D^G_kD^G_l\eta_{kl} + \sum_i\phi^A_{i0}\log(\theta_i)d^G_i \\ &+ \frac{1}{2}\sum_i\sum_{k,l}\psi_{ik}\psi_{il}\theta_i^2\phi^A_{i0}(\phi^A_{i0}\eta_{kl} - \eta_{kk}) + \sum_i\phi^A_{i0}A_{ii}\log(\frac{1}{2}\theta_i^2) \\ &+ \sum_i\phi^A_{i1}d_i\log(\widetilde{\theta}_i) + m_{bg}\log(\eta_{bg}) + \frac{1}{2}m_{bb}\log(\eta_{bb}) - \frac{1}{2}D^BD^B\eta_{bb} \\ &- D^B\cdot g\cdot\eta_{bg} + \sum_i\widetilde{\theta}_i\phi^A_{i1}(1 - \phi^A_{i1})(\eta_{bg} - \frac{1}{2}\eta_{bb}) + \sum_i\phi^A_{i1}A_{ii}\log(\frac{1}{2}\widetilde{\theta}_i^2)\end{aligned} \tag{4}$$

For the case that the observed graph has no self-loops (i.e. $A_{ii} = 0$), we obtain the simplified Eq. (14) as presented in the paper.

2.5 MLE/MAP of the parameters.

Before we solve for the MLE of the parameters we have to include the identifiability constraints $D_k^G \stackrel{!}{=} \sum_i (d_i^G + 2A_{ii})\psi_{ik}\phi_{i0}^A$ and $D^B \stackrel{!}{=} \sum_i (d_i + 2A_{ii})\phi_{i1}^A$. That is our log likelihood function has two additional terms $\mathcal{L} = \mathcal{L} - \sum_k \lambda_k (D_k^G - \sum_i (d_i^G + 2A_{ii})\phi_{ik}\phi_{i0}^A) - \lambda(D^B - \sum_i (d_i + 2A_{ii})\phi_{i1}^A)$, where λ_k and λ are Lagrangian multipliers. Similarly as above, focusing on graphs with $A_{ii} = 0$ we obtain the constraints as presented in the paper.

Since θ_i are independent of each other we can find the MLE for each of them separately. Taking the terms involving θ_i and setting the derivative to 0 we get: $\frac{\partial \mathcal{L}_{\theta_i}}{\partial \theta_i} = \frac{\phi_{i0}^A d_i^G}{\theta_i} - \sum_k \lambda_k \psi_{ik}\phi_{i0}^A$. Solving the $N + K$ system of equations we get the following solution: $\theta_i = d_i^G, \forall i$ and $\lambda_k = 1, \forall k$.

Similarly for $\widetilde{\theta}_i$ we have $\frac{\partial \mathcal{L}_{\widetilde{\theta}_i}}{\partial \widetilde{\theta}_i} = \frac{\phi_{i1}^A d_i}{\widetilde{\theta}_i} - \lambda \phi_{i1}^A$. Solving the $N + 1$ system of equations we get $\lambda = 1$ and $\widetilde{\theta}_i = d_i, \forall i$.

For the edge generating parameters we have:

$$
\begin{aligned}
\frac{\partial \mathcal{L}_{\eta_{kl}}}{\partial \eta_{kl}} &= \frac{m_{kl}}{\eta_{kl}} - D_k^G D_l^G = 0 \implies \eta_{kl} = \frac{m_{kl}}{D_k^G D_l^G} \\
\frac{\partial \mathcal{L}_{\eta_{bg}}}{\partial \eta_{bg}} &= \frac{m_{bg}}{\eta_{bg}} - D^B g = 0 \implies \eta_{bg} = \frac{m_{bg}}{D^B g} \\
\frac{\partial \mathcal{L}_{\eta_{bb}}}{\partial \eta_{bb}} &= \frac{1}{2}\frac{m_{bb}}{\eta_{bb}} - \frac{1}{2} D^B D^B = 0 \implies \eta_{bb} = \frac{m_{bb}}{D^B D^B}
\end{aligned}
$$

Next looking at the topics we get: $\frac{\partial \mathcal{L}_{t_{dk}}}{\partial t_{dk}} = \sum_i \psi_{ik}\phi_{i0}^X \left(\frac{X_{id}}{t_{dk}} - \frac{1 - X_{id}}{t_{dk} - 1} \right) = 0 \implies t_{dk} = \frac{\sum_i r_{ik} X_{id}}{R_k}$. Looking at the cluster probabilities and introducing Lagrangian multipliers to enforce $\sum_k \pi_k = 1$ we get for $\pi_k = \frac{\sum_i (1-\phi_{i3})\psi_{ik} + \alpha_k}{\sum_i (1-\phi_{i3}) + \sum_k \alpha_k}$. Finally looking at ρ_m and also introducing Lagrangian multipliers to enforce $\sum_{m=0}^4 \rho_m = 1$ we have: $\rho_m = \frac{\sum_i \phi_{im} + \beta_m}{N + \sum_m \beta_m}$

2.6 Limit case analysis for approximation

In the following we justify the simplification of the term $\mathcal{L}_{\mathcal{A}}$, by considering the limit cases when the graph grows. As we will see, both terms $\sum_i \widetilde{\theta}_i \phi_{i1}^A (1 - \phi_{i1}^A)(\eta_{bg} - \frac{1}{2}\eta_{bb})$ and $\frac{1}{2}\sum_i \sum_{k,l} \psi_{ik}\psi_{il}\theta_i^2 \phi_{i0}^A (\phi_{i0}^A \eta_{kl} - \eta_{kk})$ become neglectable.

Recap the definition of $\mathcal{L}_{\mathcal{A}}$:

$$
\begin{aligned}
\mathcal{L}_{\mathcal{A}} &= \frac{1}{2}\big(\sum_{k,l} m_{kl} log(\eta_{kl}) - D_k^G D_l^G \eta_{kl} + m_{bb}\log(\eta_{bb}) + D^B D^B \eta_{bb}\big) \\
&+ \frac{1}{2}\sum_i \sum_{k,l} \psi_{ik}\psi_{il}\theta_i^2 \phi_{i0}^A (\phi_{i0}^A \eta_{kl} - \eta_{kk}) + m_{bg}\log(\eta_{bg}) - g D^B \eta_{bg} \\
&+ \sum_i \phi_{i0}^A \log(\theta_i) d_i^G + \phi_{i1}^A \log(\widetilde{\theta}_i) d_i + \sum_i \widetilde{\theta}_i \phi_{i1}^A (1 - \phi_{i1}^A)(\eta_{bg} - \frac{1}{2}\eta_{bb})
\end{aligned}
$$

After carefully rearranging the terms η_{kl}, η_{bg}, and η_{bb}, we obtain:

$$
\begin{aligned}
\mathcal{L}_{\mathcal{A}} &= \frac{1}{2}\Big(\sum_{k,l} m_{kl} log(\eta_{kl}) - \sum_i \psi_{ik}\theta_i \phi_{i0}^A \eta_{kl} (D_l^G - \psi_{il}\theta_i \phi_{i0}^A) \\
&+ m_{bb}\log(\eta_{bb}) + \sum_i \widetilde{\theta}_i \phi_{i1}^A \eta_{bb}\big(D^B - (1 - \phi_{i1}^A)\big)\Big) \\
&+ \frac{1}{2}\sum_i \sum_k \psi_{ik}\theta_i^2 \phi_{i0}^A (-\eta_{kk}) + m_{bg}\log(\eta_{bg}) - \sum_i \widetilde{\theta}_i \phi_{i1}^A \eta_{bg}\big(g - (1 - \phi_{i1}^A)\big) \\
&+ \sum_i \phi_{i0}^A \log(\theta_i) d_i^G + \phi_{i1}^A \log(\widetilde{\theta}_i) d_i)
\end{aligned}
$$

Now, splitting the $\sum_{k,l}$ into $\sum_{k \neq l}$ and $\sum_k$, we can also rearrange the terms η_{kk}. Thus, we obtain:

$$
\begin{aligned}
\mathcal{L}_{\mathcal{A}} &= \frac{1}{2}\Big(\sum_{k \neq l} m_{kl} log(\eta_{kl}) - \sum_i \psi_{ik}\theta_i \phi_{i0}^A \eta_{kl} (D_l^G - \psi_{il}\theta_i \phi_{i0}^A) \\
&+ \sum_k m_{kk} log(\eta_{kk}) - \sum_i \psi_{ik}\theta_i \phi_{i0}^A \eta_{kk} (D_k^G - \psi_{ik}\theta_i \phi_{i0}^A + \theta_i) \\
&+ m_{bb}\log(\eta_{bb}) + \sum_i \widetilde{\theta}_i \phi_{i1}^A \eta_{bb}\big(D^B - (1 - \phi_{i1}^A)\big)\Big) \\
&+ m_{bg}\log(\eta_{bg}) - \sum_i \widetilde{\theta}_i \phi_{i1}^A \eta_{bg}\big(g - (1 - \phi_{i1}^A)\big) \\
&+ \sum_i \phi_{i0}^A \log(\theta_i) d_i^G + \phi_{i1}^A \log(\widetilde{\theta}_i) d_i)
\end{aligned}
$$

Which can further be written as:

$$
\begin{aligned}
\mathcal{L}_{\mathcal{A}} &= \frac{1}{2}\Big(\sum_{k \neq l} m_{kl} log(\eta_{kl}) - \sum_i \psi_{ik}\theta_i \phi_{i0}^A \eta_{kl} D_l^G (1 - \frac{\psi_{il}\theta_i \phi_{i0}^A}{D_l^G}) \\
&+ \sum_k m_{kk} log(\eta_{kk}) - \sum_i \psi_{ik}\theta_i \phi_{i0}^A \eta_{kk} D_k^G (1 - \frac{\psi_{ik}\theta_i \phi_{i0}^A}{D_k^G} + \frac{\theta_i}{D_k^G}) \\
&+ m_{bb}\log(\eta_{bb}) + \sum_i \widetilde{\theta}_i \phi_{i1}^A \eta_{bb} D^B (1 - \frac{(1 - \phi_{i1}^A)}{D^B})\Big) \\
&+ m_{bg}\log(\eta_{bg}) - \sum_i \widetilde{\theta}_i \phi_{i1}^A \eta_{bg} g (1 - \frac{(1 - \phi_{i1}^A)}{g}) \\
&+ \sum_i \phi_{i0}^A \log(\theta_i) d_i^G + \phi_{i1}^A \log(\widetilde{\theta}_i) d_i)
\end{aligned}
$$

Introducing the following abbreviation $a_{il} = \psi_{il}\theta_i\phi_{i0}^A$ and noticing the $1 - \phi_{i1}^A = \phi_{i0}^A$, we finally arrive at

$$
\begin{aligned}
\mathcal{L}_{\mathcal{A}} = \frac{1}{2}\Big(& \sum_{k \neq l} m_{kl} log(\eta_{kl}) - \sum_i \psi_{ik}\theta_i\phi_{i0}^A\eta_{kl}D_l^G \underbrace{(1 - \frac{a_{il}}{D_l^G})}_{(1)} \\
& + \sum_k m_{kk} log(\eta_{kk}) - \sum_i \psi_{ik}\theta_i\phi_{i0}^A\eta_{kk}D_k^G \underbrace{(1 - \frac{a_{ik}}{D_k^G} + \frac{\theta_i}{D_k^G})}_{(2)} \\
& + m_{bb}\log(\eta_{bb}) + \sum_i \widetilde{\theta}_i\phi_{i1}^A D^B\eta_{bb} \underbrace{(1 - \frac{\phi_{i0}^A}{D^B})}_{(3)} \Big) \\
& + m_{bg}\log(\eta_{bg}) - \sum_i \widetilde{\theta}_i\phi_{i1}^A\eta_{bg}g \underbrace{(1 - \frac{\phi_{i0}^A}{g})}_{(4)} \\
& + \sum_i \phi_{i0}^A \log(\theta_i)d_i^G + \phi_{i1}^A \log(\widetilde{\theta}_i)d_i)
\end{aligned}
$$

For the limit case it is now sufficient to consider the terms (1)-(4). Note that if we replace each of the terms in the underbraces (1) through (4) with the value of 1 we obtain the simplified equation in the paper. Thus, if we can show that (1)-(4) converge to 1 in the limit case, the approximation error becomes neglectable. Indeed, as we will see this holds for almost all the cases – and in the cases where it does not hold, the two terms $\sum_i \widetilde{\theta}_i\phi_{i1}^A(1 - \phi_{i1}^A)(\eta_{bg} - \frac{1}{2}\eta_{bb})$ and $\frac{1}{2}\sum_i\sum_{k,l}\psi_{ik}\psi_{il}\theta_i^2\phi_{i0}^A(\phi_{i0}^A\eta_{kl} - \eta_{kk})$ vanish due to another reason.

We start with the simplest case (4): Since $g = \sum_{i=1}^N \phi_{i0}^A$, the term (4) obviously approaches 1. More formally, we can distinguish two cases: First, if $\lim_{N\to\infty} g = \infty$, then clearly $(\frac{\phi_{i0}^A}{g}) \to 0$ since ϕ_{i0}^A is bounded by 1 and the denominator grows faster. Second, if $\lim_{N\to\infty} g = c$ for some constant c, the series converges. Thus, it has to hold $\lim_{i\to\infty}\phi_{i0}^A = 0$. And therefore again $(\frac{\phi_{i0}^A}{g}) \to 0$.[1]

The exactly same argumentation holds for the term (1): We have $D_l^G = \sum_{i=1}^N a_{il}$. Either it holds $\lim_{N\to\infty} D_l^G = \infty$, in which case, in the fraction the denominator grows faster than the nominator, i.e. the fraction becomes 0. Or we have $\lim_{N\to\infty} D_l^G = c$ for some constant c. The series converges and, thus, $\lim_{i\to\infty} a_{il} = 0$. In this case the nominator approaches 0. In both cases the fraction approaches 0 and therefore (1) approaches 1.

Let us now consider the term (3): Recap that $D^B = \sum_{i=1}^N \widetilde{\theta}_i\phi_{i1}^A$. In the first case, if $\lim_{N\to\infty} D^B = \infty$ then (3) approaches 1. Note again, that this result combined with the above result for (4) means that we can safely drop the term $\sum_i \widetilde{\theta}_i\phi_{i1}^A(1 - \phi_{i1}^A)(\eta_{bg} - \frac{1}{2}\eta_{bb})$. In the second case, $\lim_{N\to\infty} D^B = c$ for some constant c. Then $\lim_{i\to\infty}\widetilde{\theta}_i\phi_{i1}^A = 0$ has to hold. Accordingly, almost all terms in $\sum_i \widetilde{\theta}_i\phi_{i1}^A(1-\phi_{i1}^A)(\eta_{bg}-$

[1] Note that in general the following holds: If $\lim_{N\to\infty}\sum_{i=1}^N x_i = c$ for some constant c (i.e. the series converges), then $\lim_{i\to\infty} x_i = 0$ has to hold.

$\frac{1}{2}\eta_{bb})$ evaluate to zero (note that the variables are all bounded by 1). We can again safely drop the term.

Finally consider the term (2): Recap that $D_k^G = \sum_{i=1}^N a_{ik} = \psi_{ik}\theta_i\phi_{i0}^A$. Again, for the case $\lim_{N\to\infty} D_k^G = \infty$ the overall term clearly approaches 1. Second, for the case $\lim_{N\to\infty} D_k^G = c$, the terms a_{ik} approach 0. That is, the same argumentation as for (3)+(4) can be used: in combination with the result for (1), the term $\frac{1}{2}\sum_i \sum_{k,l} \psi_{ik}\psi_{il}\theta_i^2\phi_{i0}^A(\phi_{i0}^A\eta_{kl} - \eta_{kk})$ can be dropped since almost all elements evaluate to zero.

Overall, in all cases (i.e. independent whether the individual series converge or diverge), the resulting error we make by the approximation approaches zero. Finally note that is not possible that all series converge at the same time. At least one of the series has to diverge when the graph grows since either infinitely many good or infinitely many anomalous nodes have to be added (or both). This means that the overall term $\mathcal{L}_\mathcal{A}$ in the simplified version has to diverge as well. Thus, while this term grows, the error gets smaller.

References

[1] C. M. Bishop. Pattern recognition. *Machine Learning*, 128, 2006.

[2] B. Perozzi, L. Akoglu, P. I. Sánchez, and E. Müller. Focused clustering and outlier detection in large attributed graphs. In *ACM SIGKDD*, pages 1346–1355, 2014.