# Deriving the Political Affinity of Twitter Users from Their Followers

Giorgos Stamatelatos, Sotirios Gyftopoulos, George Drosatos and Pavlos S. Efraimidis
Dept. of Electrical and Computer Engineering, Democritus University of Thrace, Kimmeria, GR67100 Xanthi, Greece
{*gstamat,sgyftopo,gdrosato,pefraimi*}*@ee.duth.gr*

***Abstract*—In this work, we show that Twitter users can reveal valuable political information about particular Nodes of Interest (NOIs) they opt to follow. More precisely, we utilize an interesting graph projection method and a series of algorithmic approaches, such as modularity clustering, a minimum linear arrangement (MinLA) approximation algorithm and the DeGroot opinion update model in order to reveal the political affinity of selected NOIs. Our methods, which are purely structure-based, are applied to a snapshot of the Twitter network based on the user accounts of NOIs, consisting of the members of the current Greek Parliament along with their respective followers. The findings confirm that the information obtained can portray with significant precision the political affinity of the NOIs. We, furthermore, argue that these methods are of general interest for imprinting the political leaning of other NOIs, for example news media, and potentially classifying them in respect to their political bias.**

***Index Terms*—Social Network Analysis, Twitter Followers, Political Affinity**

## I. Introduction

Twitter, an online news and social networking service, has been subject of scientific research for at least a decade. Users in Twitter can follow other users in order to receive short messages posted by them, which are called tweets. The follower relationships of Twitter naturally convey an inherent directed graph structure, where vertices are the user accounts and edges represent the follower-to-followee relationship. The interpretation of these links vary across contexts: they may represent intimate relationships, common interests, an intent in news briefing and many others.

The significance of Twitter in research is partially because it supplies means to comprehend social relationships and influence dynamics in human societies. Various studies examine the structural and topological characteristics of the Twitter network, for example in [1] and [2] the authors study concepts related to user influence and centrality. Furthermore, the Twitter network has been previously used for stock market predictions [3], trend detection [4], geo-locating users [5] and more.

A distinct characteristic of Twitter is the presence of politically related actors, for example politicians or other party representatives, public officials, candidates as well as news media and news sources. These actors engage in the social platform as part of their political campaigns or utilize it as a means of political deliberation and advocacy. The topic of political deliberation in social networks is relevant and has received a substantial attention, especially through its applications to the identification of political bias in news sources.

In this work we study the topic of deriving the political affinity of particular users of interest (NOIs) by using the structural features of the Twitter network. The methods we propose rely only on the social ties formed among relevant parties in the network. More specifically, we consider the NOIs to be the members of the current Greek Parliament (MPs), although the set of NOIs can be enriched with other politically engaged actors as well.

Overall, our approach relies on the assumption that people's political preferences will, on average, reflect those of the politicians they follow. Prior literature on *selective exposure* to political information suggests this assumption is reasonable since people seek after information from those with similar political views [6]. Following decisions are considered signals that provide information about Twitter users' perceptions of both their ideological location and that of the political accounts [7].

The dataset used in the experiments, described in Section III, contains the social ties formed among the NOIs with their followers. In Section IV we demonstrate the existence of rich political information within the dataset by evaluating the effectiveness of a well-established community detection approach and, moreover, lay out the experimentation settings. In Sections V and VI two different methods to extract political affinity information from the dataset are presented, the minimum linear arrangement problem (MinLA) and the DeGroot opinion update model with stubborn agents, respectively. Finally, Section VII concludes this paper and presents suggestions for future work.

## II. Related Work

A number of previous studies have promoted concepts related to the detection and analysis of political affinity. In [8], the authors show that Twitter is used extensively for political deliberation and evaluate whether tweets reflect the current offline political sentiment. In [9], the values of user attributes such as political orientation or ethnicity are inferred, while in [10] an example application to determine political leanings from tweets is demonstrated. These methods operate by examining the content of tweets. The approaches presented in this paper utilize algorithms that only rely on the topological and structural characteristics of the Twitter network. A fundamental principle is to utilize the follower set of a node to characterize their political affinity, rather than the content of their own tweets or other information that directly originates from their own actions. An additional effect is that users of importance cannot easily handpick their followers, who adapt to the online and offline political scene.

Furthermore, in [11], the authors construct the politician-journalist graph and attain multiple conclusions regarding the network structure. Moreover, the study of [12] is the identification of the characteristics of political parties and the political leaning of users in social media. The data scheme used in these reports is similar to the one used here but our focus and methodology are distinct.

An interesting recent study in [13] attempts to infer the political leaning of news outlets in the US by characterizing the followers and then relaying the followers' preference to the news outlets that they opt to follow. The authors claim that, overall, users tend to follow politicians with similar views and that those who follow Congresspeople on Twitter may have more polarized political tendencies that the overall US population. The results are achieved using the American for Democratic Actions (ADA) scores. The objectives of our work are similar to those in [13] but in this paper we establish methods that work in a multi-party context which, moreover, don't require a quantitative starting point, like the ADA scores.

In [7], the author uses the structural characteristics of the Twitter network to extract the political positions of politicians, users and news sources in five countries. He proposes a *Bayesian spatial following model of ideology* based on the popularity of the politicians, the political interest of users and their estimated ideal points on the political spectrum in order to predict the probability of a user following a politician. Although the author's hypothesis coincides with our hypothesis (i.e., the mere structure of the Twitter network suffices for the extraction of the political inclination of specific users), his proposed model applies extensive filtering on the users' dataset (e.g., geolocation, tweet activity, number of followers) while in our approach we use raw data for our algorithms without filtering and without any other knowledge of the users' characteristics.

Overall, our approach relies only on the social ties formed among the NOIs and their followers. The methods proposed in this work utilize the nodes of the implicit graph structure simply as their Twitter IDs, and no additional knowledge about these user accounts is required. Furthermore, our methods are easily reproducible and can be implemented without complex filtering or preprocessing. As shown, prior literature suggests that the study of political leaning identification in the Twitter follower network is possible. Here we provide new evidence based on novel algorithmic approaches that are not commonly discussed in this context, given that prior work on this subject is fragmented.

## III. The Twitter Follower Dataset

The dataset that we assemble and use in this work is based on the Twitter accounts of members of the current Greek Parliament (MPs); we refer to these actors as *Nodes of Interest* (NOIs). The set of NOIs was acquired from the official website of the Greek Parliament[1] without discrimination. Summarily, the list of NOIs consists of 300 MPs, of which 166 have a Twitter account. As a result, 134 MPs are missing from the dataset. Among the disregarded MPs is the party KKE, one of the eight political parties, representing the left wing of the Greek Parliament, of which none of the MPs had a Twitter account. A breakdown of the NOIs is presented in Table I, showing the number of actors with a Twitter account present in the dataset along with the total number that we obtained.

TABLE I: Breakdown of NOIs into parties.

| Group | Dataset | Acquired |
|---|---|---|
| SYRIZA | 62 | 145 |
| ND | 61 | 76 |
| DHSY | 17 | 20 |
| XA | 10 | 16 |
| POTAMI | 6 | 6 |
| EK | 2 | 6 |
| KKE | 0 | 15 |
| ANEL | 4 | 9 |
| Independent | 4 | 7 |
| Totals | 166 | 300 |

We complete our dataset by crawling the followers of the NOIs using the Twitter API to construct a bipartite graph of 162 NOIs, 740,537 followers and 2,402,237 edges. An edge $E_{ij}$ between a NOI $i$ and a follower $j$ exists iff $j$ is following $i$. The dataset was constructed on April 2018. It is worth mentioning that we only considered the 162 NOIs with a party affiliation (and not the 4 independent MPs shown in Table I); the intuition behind this decision is that users that don't exhibit political engagement do not convey significant information about the objectives of this work.

The dataset, along with other supplementary material about this work, is available online[2].

## IV. The Overlap Projection

A necessary step for deriving the political affinity from the Twitter follower dataset is to initially confirm the hypothesis that Twitter followers can reveal the political affinity of NOIs. Thus, in this section, we apply community detection and show that it can reveal the underlying political structure of our dataset.

The dataset, however, is massive and possibly incompatible with generic graph processing algorithms due to its bipartite nature. Thus, we transform the graph to its one-mode projection onto the NOIs, an extensively used method for compressing information about bipartite networks [14]. The

[1] https://www.hellenicparliament.gr/en/Vouleftes/Ana-Koinovouleftiki-Omada/

[2] https://euclid.ee.duth.gr/publication/socialcom-2018/

one-mode projection of a bipartite network $G = (X, Y, E)$ onto $X$ ($X$ projection for short) is a weighted unipartite network $G' = (X, E')$ containing only $X$ nodes, where $weight(E'_{ij}) = \beta_G(X_i, X_j)$. Many real world networks are naturally modeled as bipartite graphs, especially in social systems, like the Twitter follower network we use in this work. The projection allows simplification of the network and compatibility with unipartite algorithms by selecting a function $\beta_G$ that doesn't incur a significant deficit of information about the original network.

The purpose of this experiment is to identify a similarity measure $\beta_G$ that preserves the quality of the information inherent in the dataset. We have compared multiple similarity functions commonly found in the literature, namely the Overlap coefficient $h_G$, the Pearson correlation coefficient, the Cosine similarity, the Jaccard index and the F1 score (Sorensen-Dice coefficient). We were unable to include more complex measures due to the large scale of the input data, for example the original SimRank [15] algorithm has a space requirement of $O(n^2)$. Our observations, however, suggest little room for further improvement over existing effectiveness. These measures were applied on the follower sets of the NOIs, so that the weight of the edge incident to NOIs $i$ and $j$ becomes a function of their followers' sets, i.e. $weight(E'_{ij}) = \beta_G(adj(X_i), adj(X_j))$, where $adj(Z)$ denotes the set of nodes that are adjacent to $Z$ (their followers).

Our experiments indicate that, for this context, the Overlap coefficient is strictly better and achieves consistently positive results. The Overlap coefficient (also known as Simpson coefficient) is a modification of the Jaccard similarity index:

$$h_G(X_i, X_j) = \frac{|adj(X_i) \cap adj(X_j)|}{min(|adj(X_i)|, |adj(X_j)|)}.$$

The measure has appeared in a number of articles, for example in [16] for the purpose of studying affiliation networks and in [17] for text mining applications. Overlap coefficient expresses a form of similarity between two nodes; it assumes the value of 1 if the nodes are identical and a value of 0 if they have no common follower.

The evaluation of these measures was performed using a community detection approach on each of the 5 resulting projections, considering that the partition of MPs in political parties is known. The participation of a MP in their political party (Table I) is an objective indication about their political affinity and constitutes a reference in this clustering scenario. The aim of community detection in graphs is to identify the modules and, possibly, their hierarchical organization, by using only the information encoded in the graph topology [18]. More specifically, we use the algorithm in [19] on the different projections of our graph, a heuristic method that is based on modularity optimization and is commonly known as *Louvain optimization*. The algorithm is well established and has seen extensive use in the field of social networks [20]. The *resolution* parameter of the algorithm determines the number of communities in the partition. Its complexity is linear on typical and sparse data.

We apply the clustering algorithm on each projection to partition the MP vertex set into disjoint groups and evaluate each partition using different measures with the concept underlined in [21, Section 2.2.1]. Specifically, the evaluation of the projection methods is performed using the Jaccard index, the Simple Matching Coefficient (SMC), the F1 score, the Normalized Mutual Information (NMI), the Pearson correlation and the Cosine similarity. These measures are used to assess the effectiveness of a partition and are different from the measures used to construct the projection, although some of them are used for both purposes. The results are shown in Table II. The rows of the table refer to the similarity functions used for the projection. For brevity, each function is represented only by the resolution with the best evaluation. For comparison, the random partition is also included in the table. The columns denote the evaluation measures.

TABLE II: Clustering evaluation of various projections.

| **$\beta_G$/Resol.** | **Evaluation measure** | | | | | |
|---|---|---|---|---|---|---|
| | Jaccard | SMC | F1 | NMI | Pearson | Cosine |
| $h_G$/0.95 | **.876** | **.959** | **.934** | **.733** | **.904** | **.934** |
| Pearson/1.00 | .549 | .836 | .709 | .289 | .598 | .710 |
| Cosine/0.90 | .542 | .839 | .703 | .296 | .600 | .707 |
| Jaccard/0.90 | .536 | .834 | .698 | .284 | .590 | .701 |
| F1/0.90 | .526 | .830 | .689 | .273 | .579 | .693 |
| Random | .107 | .642 | .194 | .000 | .000 | .208 |

The results indicate that the overlap coefficient firmly outperforms other functions commonly mentioned in the literature on all evaluation measures. Its significance is also verified by the experiments in Sections V and VI. It is, furthermore, evident that the modularity-based clustering on the overlap projection produces a prominent partition, indicating the existence of rich political affinity information within the Twitter follower network.

Figure 1 displays a force-directed visualization, produced by Gephi [22], of the MP projection using the $h_G$ function. Vertices are colored by their modularity class, with a resolution of 0.95. The respective party of each node (the ground truth) is given as a text label within the node while the party IDs are given in the legend in the top left corner. An important visual observation is that the accuracy of the identified clusters is remarkably high, which coincides with the results in Table II. SYRIZA and ND are clearly identified with the respective clusters almost flawlessly, and for the XA party, there is not a single error. In this particular visualization, we have chosen a resolution parameter of the community detection algorithm which gives five classes of nodes. Consequently, the smaller parties are grouped into a single cluster. Other values for the resolution parameter gave larger number of classes, but the overall accuracy was worse. This aspect, i.e., the optimal selection of the resolution parameter, needs to be further examined. However, since the number of parties in the Parliament and the number of members in each party are known, the parameter tuning is a feasible task. Additionally, the layout provides a visual perception of the close association between modularity clustering and force-directed placement [23].

In the rest of the document, we refer to the overlap projection of the Twitter follower network onto the NOIs as *NOI projection* $G$. In our study, due to reasons stated in this section, we don't use the original graph and experiment exclusively in $G$. The dataset link provided in Section III includes this projection.

## V. MINIMUM LINEAR ARRANGEMENT

In this section, we introduce the *Minimum Linear Arrangement* (MinLA) problem and apply a solution on $G$ for the purpose of revealing the clustering features of our dataset. We argue that MinLA is suitable for studying the dataset and additionally justify the effectiveness of the overlap coefficient on a completely different setting than typical social network analysis tools. To the best of our knowledge the MinLA problem has not been applied on a social network theme in prior literature.

The *Minimum Linear Arrangement* (MinLA) problem consists in finding an ordering of the nodes of a weighted graph, such that the sum of the weighted edge lengths is minimized. More formally, given a finite graph $G = (V,E)$ of order $n$ with weighted adjacency matrix $w$, the MinLA problem is the problem of finding a vertex labeling $f \to \{1,2,...,n\}$ such that the sum $\sum_{(u,v)\in E} w_{uv}|f(u) - f(v)|$ is minimized over all possible labelings [24]. On general graphs, MinLA is an NP-Complete problem, thus one has to resort to heuristics or approximation algorithms to obtain a solution.

There are no known approximation algorithms that achieve "good" approximation guarantees for the MinLA problem. The best known result, is the $O(\sqrt{\log n}\log\log n)$-approximation algorithm presented in [25]. On the other hand, in [26] it is shown that no Polynomial Time Approximation Scheme (PTAS) exists for MinLA and in [27] that it is SSE-hard to approximate MinLA to any fixed constant factor.

For this experiment, we implemented a simplistic randomized local search algorithm, repeated over a set of uniformly random initial rankings, which approximately leads to the minimum LA. Given an initial guess of the arrangement we perform a sequential series of steps to determine a local minimum of the cost function, the *fast converge phase* and the *local converge phase*. During the fast phase, the algorithm performs random swaps on the elements of the arrangement for a number of repetitions, and maintains the best arrangement in terms of cost. The purpose of this phase is to allow the algorithm to quickly descend close to a local minimum while the number of repetitions involved determine the convergence rate. We selected to perform this step $n^2$ times as we empirically observed a sufficiently quick convergence for this setting. During the local phase we validate that the current LA is the local minimal cost LA by performing all possible swaps in it; if there is a swap that improves the cost we restart the process until we identify the local minimum. The above process of computing a local minimum is repeated several times with random initial arrangements and the best solution is kept. The scheme is presented in Algorithm 1.

**Algorithm 1:** Local Search MinLA algorithm

```
Function localMin (a: Array)
    for n^2 times do
        Perform a random swap on a to create a'
        If it reduces the cost set a ← a'
    end
    for x in [1,n), y in (x,n] do
        Perform the swap (x,y) on a to create a'
        If it reduces the cost set a ← a' and restart the
          loop
    end
end

Function main (a: Array, reps: Int)
    for reps times do
        Shuffle a to create a' and invoke localMin(a')
        If the cost of a' is lower than a set a ← a'
    end
end
```

Initially, we apply this algorithm to the NOI projection graph to obtain the (approximately) minimum cost arrangement $m$, with a cost of $C(m) = 162.2K$. Furthermore, because in this experiment we deal with ordinal data, we also define the concept of party ordering. Our dataset contains 7 parties, so there are $\rho = 7! = 5{,}040$ possible orderings. Each of these orderings can be flattened to a ranked list of MPs, where MPs of the same party are tied on the same rank. Thus, there are also $\rho$ flattened MP ranked lists denoted as $R_i$, $1 \leq i \leq \rho$.

We assess the correlation of $m$ with every ordering $R_i$ using the Kendall tau-b ($\tau_B$) correlation coefficient [28], which is a statistic used to measure the ordinal association between two measured quantities. The tau-b correlation coefficient is a generalization of the Kendall tau-a coefficient that accounts for ties in the input lists, specifically present in the $R_i$ orderings. It is worth noting that Kendall tau-b is in range $[-1,1]$ but, since in our context the LAs cannot contain ties, the maximum value is

$$K_{max} = \frac{T(n) - \sum_i T(t_i)}{\sqrt{T(n)}\sqrt{T(n) - \sum_i T(t_i)}}, \text{ where } T(x) = \frac{x(x-1)}{2},$$

which equals 0.8361 because $t = [62,61,17,10,6,4,2]$ (Table I). Afterwards, we find the party ordering with the highest correlation to $m$, defined as $R_q$ where

$$q = \underset{i\in[1,\rho]}{\arg\max}\, \tau_B(m,R_i).$$

More specifically, $R_q$ is [SYRIZA, ANEL, POTAMI, DHSY, ND, EK, XA] of which the correlation with $m$ is remarkable with $\tau_B(m,R_q) = 0.7529$ (90% of the maximum), proving that the MinLA problem definition can reveal the clustering features of our dataset. For comparison, it can be shown that

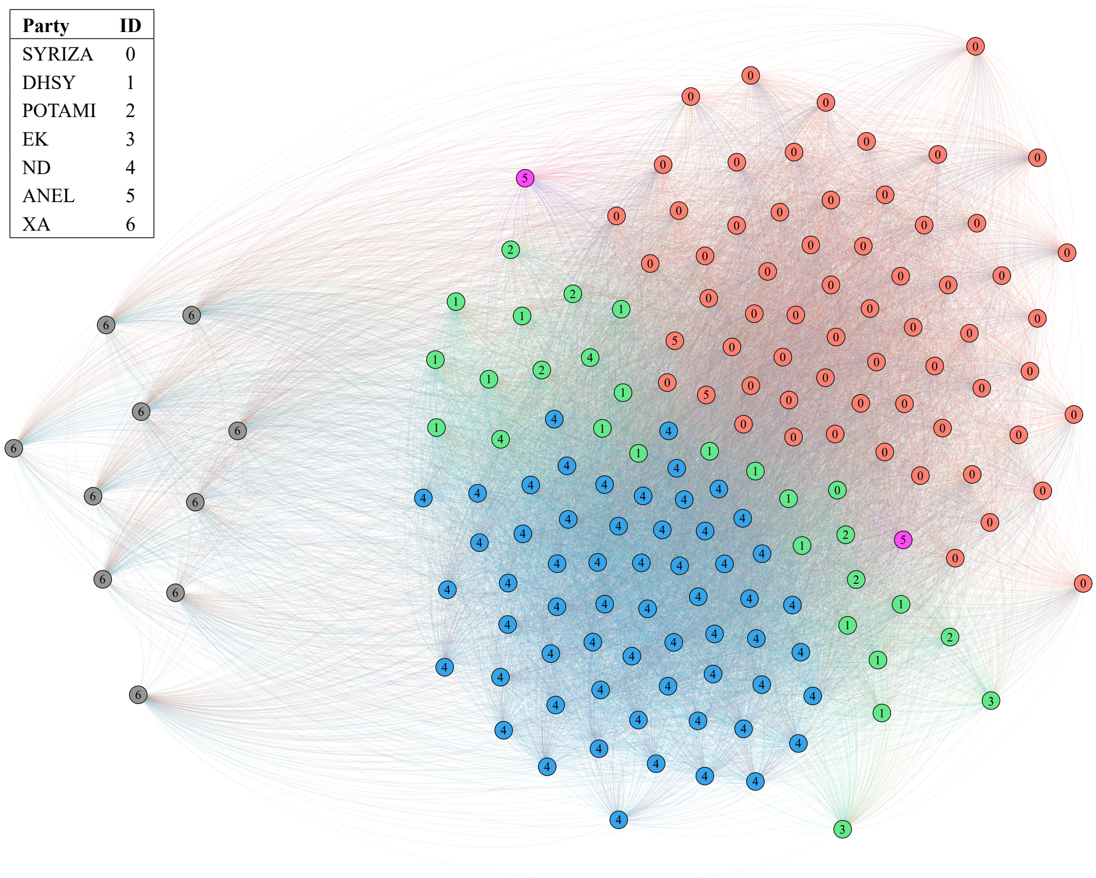


Fig. 1: Force-directed visualization of the NOI projection.

the average distance of two nodes in a random LA is $(n+1)/3$, which is stated in the following lemma.

**Lemma 1.** *The average distance of two nodes in a random LA is* $(n+1)/3$.

*Proof.* Let $X$ and $Y$ be two random variables for the positions of the two nodes, respectively, in the LA. First, assume $X<Y$. Then, the following sum $S_1$ is:

$$\sum_{x=1}^{n}\sum_{y=x+1}^{n} P[X=x]\cdot P[Y=y|X=x]\cdot(y-x)$$

$$=\frac{1}{n}\frac{1}{n-1}\sum_{x=1}^{n}\sum_{y=x+1}^{n}(y-x)=\frac{1}{2n(n-1)}\sum_{x=1}^{n}(n-x)(n-x+1)$$

Assuming $X>Y$, the corresponding sum $S_2$ has the same value $S_2=S_1$. The average distance is equal to the sum $S_1+S_2$. Adding $S_1$ and $S_2$, and then simplifying gives $(n+1)/3$. ■

Thus, the cost of the random LA is $\sum_{ij} w_{ij}(n+1)/3$, which equals $225.1K$ in our projection. Figure 2 shows the relation between the MinLA cost $C(x)$ and the tau-b correlation $\tau_B(x,R_q)$ of the current minimum cost LA $x$ as the algorithm converges. The south-most point in this figure is $m$. This figure also shows that there is a very strong, almost linear relationship of the LA cost function with the correlation function.

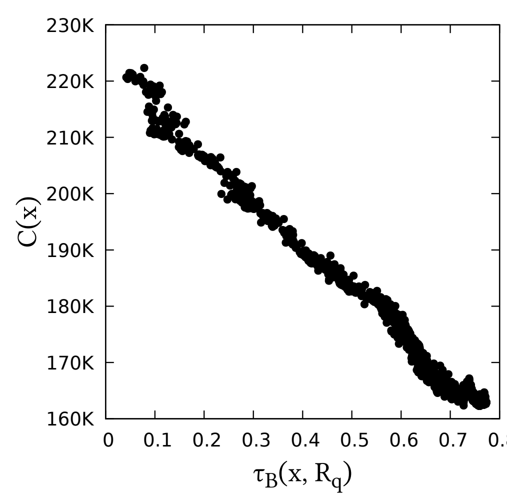


Fig. 2: Convergence of the arrangement cost as a function of the correlation with $R_q$.

Furthermore, our results are strengthened by the experiments conducted for all possible party orderings $R_i$, which is displayed as a scatter plot in Figure 3. There are $\rho$ points in the plot and each one represents a single permutation $R_y$. Each $R_y$ is associated to its correlations with $R_q$ and $m$. The relationship

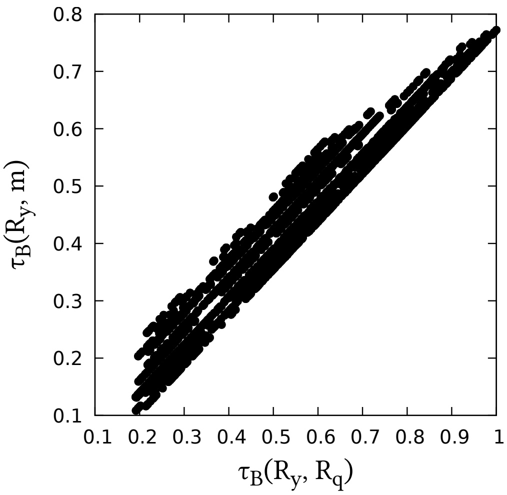


Fig. 3: Scatter plot of party permutations displaying their respective correlations with $R_q$ and $m$.

highlights the significance of the MinLA application on $G$ by displaying the wide range of correlations that any ordering $R_y$ has with $m$ or $R_q$. Moreover, the plot shows that the low-cost linear arrangement $m$ strongly favors the specific party ordering $R_q$. The more correlated an ordering $R_y$ is to $R_q$, the more correlated it is, in an almost linear way, to the whole arrangement $m$. We consider this further evidence that the MinLA problem is suitable for application on this context, and possibly constitutes an algorithmic tool for the analysis and understanding of social networks.

A comparison of $R_q$ to the arrangement of the parties based on their ideological identity reveals interesting properties of our result and of the inherent political information in our dataset. According to their self-identification and data from additional resources (e.g. Wikipedia), the most credible arrangement of the parties on the left-to-right political spectrum is [SYRIZA, DHSY, POTAMI, EK, ND, ANEL, XA]. In $R_q$, ANEL is adjacently positioned to SYRIZA, an oxymoron phenomenon that can be justified by the fact that ANEL and SYRIZA are in governmental coalition and, thus, their ties are strong in the Twitter follower dataset. Furthermore, the swap of DHSY and POTAMI in $R_q$ is inconsequential, especially after their recent deliberations (in April 2018) about the formation of a new upcoming coalitional party (KINAL) for the upcoming elections. The misplacement of EK can be attributed to its small footprint (2 MPs) and, hence, by deficient information. In general, we can argue that $R_q$ outlines the parties on one dimension according to the followers' criteria that are a combination of the left-to-right political perspective and the pro and anti-government feeling.

## VI. The DeGroot Model Approach

Twitter is a prominent social network that is mainly used for the dissemination of political information and beliefs; it is only natural to utilize a prominent model of opinion diffusion in order to evaluate the political information in our dataset.

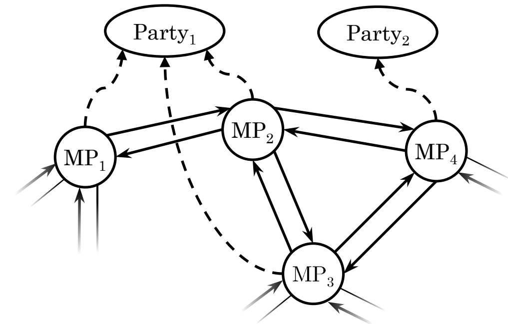


Fig. 4: Abstract example of the NOI projection $G$ enriched with party nodes.

Moris H. DeGroot presented in [29] a simple yet efficient model about opinion diffusion in a social graph. The core idea of his model is that individuals tend to adopt the opinions of their friends. According to the model's *opinion update rule*, given a social graph $G = (V, E, O)$, where $V$ represents the vertices (i.e., individuals), $E$ the edges (i.e., friendships) amongst them and node $i \in V$ has a real valued opinion $o_i$ about a specific topic, each individual updates $o_i$ to $o_i'$ by averaging the opinions of its friends. When *trust factors* are introduced to the friendships (i.e., weights), each member updates its opinion according to the weighted average of its friends' opinions. The process is repeated and DeGroot studied its convergence. Furthermore, he stated the conditions under which the social graph reaches a consensus and underlined the mathematical coherence of the process to Markov chains. He proved that the final opinion, when convergence occurs, depends solely on the structure of the graph and the initial opinions of its members.

In [30], Ghaderi and Srikant enriched DeGroot's model with *stubborn agents* (i.e. nodes that are fully or partially biased towards an opinion) and studied its convergence. They remarked the common underpinnings of their extension with Markov chains and proved that "the model converges to a unique equilibrium where the opinion of each agent is a convex combination of the initial opinions of the stubborn agents". Moreover, the contribution of stubborn agent $s$ in node's $i$ final opinion is the probability of a random walk hitting $s$ given it started from $i$ (*hitting probability*).

In our work, we utilize the findings of Ghaderi and Srikant to evaluate the political information in our dataset. We consider the NOI projection $G$ of the Twitter follower network as a social graph where the weights of the links correspond to the trust factors of the nodes to their neighbors. The undirected edges of the graph were duplicated into two opposite directed arcs and, in order to abide by the DeGroot model restrictions, the weights of each node's outgoing arcs were normalized to unity. Seven additional nodes were introduced to the graph that represented the political parties and each MP's node was linked to the corresponding party.

TABLE III: Leave-one-out cross-validation results.

| Party | Size | Hits | Misses | Leaks |
|---|---|---|---|---|
| SYRIZA | 62 | 53 | 9 | 9:ANEL |
| ND | 61 | 59 | 2 | 2:POTAMI |
| DHSY | 17 | 16 | 1 | 1:POTAMI |
| XA | 10 | 10 | 0 | - |
| POTAMI | 6 | 1 | 5 | 5:DHSY |
| ANEL | 4 | 3 | 1 | 1:SYRIZA |
| EK | 2 | 0 | 2 | 1:POTAMI, 1:ANEL |
| Total | 162 | 142 | 20 | |
| | | 87.7% | 12.3% | |

Figure 4 presents an abstract example of the enriched NOI projection $G$ with four MP nodes ($MP_1$, ..., $MP_4$) and two party nodes ($Party_1$, $Party_2$). When a MP updates its opinion according to the DeGroot model, its friendships to other MPs (outgoing compact arcs) are used while in the case when the MP is converted to a stubborn agent, the link to its party (dashed arc) is utilized. Hence, the direct link of a MP to its party and the MP's friendship are mutually exclusive.

In order to evaluate the political information in our graph, we resorted to the the leave-one-out cross-validation method. We conducted a series of experiments where in each case the directed arc of a selected MP to its party was ignored while the rest of MPs were influenced solely by their party, thus, transforming them into stubborn agents. The selected MP's friendships with other NOIs were used to calculate a random walk's hitting probabilities to every party's node given it originated by the MP's node. The experiments were implemented using PRISM [31], a tool that is widely used to analyze models that exhibit probabilistic behavior (e.g. Markov chains). Since the parliamentary groups are uneven, the evaluated probabilities were divided by the corresponding group's size in order to compute the *uniform per party influence* and avoid any dominance effect by the parties with large parliamentary groups. The uniform influences were used to classify each MP to a party based on their random walk. A hit was considered when the party with the greatest uniform influence on the MP coincided with its actual party.

The results presented in Table III denote that our approach achieves surprisingly high hit rate (87.7%), a clear indication that our dataset contains significant political information. Interesting observations can be pointed out by the leaks that form the miss rate (12.3%). In the case of SYRIZA, all leaked MPs (9) are misclassified to ANEL (i.e., SYRIZA's governmental partner) while the single leaked MP of ANEL is assigned to SYRIZA. Furthermore, all leaked MPs of POTAMI (5) are assigned to DHSY while DHSY's single leaked MP is appointed to POTAMI, a phenomenon that is in accordance with their strong political ties considering their recent deliberations about the formation of a unified party.

The results of the DeGroot model approach are in agreement with the findings of the MinLA algorithm. The adjacent positions of SYRIZA and ANEL in $R_q$ are highlighted by the misclassification of SYRIZA's MPs to ANEL and vice versa in the DeGroot experiments. The same phenomenon can also be noted with the MPs of DHSY and POTAMI. The common findings of the two algorithms from different scientific fields confirm the validity of our argument that Twitter users can reveal valuable political information about specific NOIs.

## VII. Conclusions

The purpose of this work is to study and assess the possibility of deriving political affinity of particular nodes of interest using the Twitter follower network. Our results suggest additional evidence about the validity of the hypothesis that Twitter followers can portray the political leanings of their followees. Our approach is simple to implement and easy to reproduce while delivering very significant accuracy. Furthermore, the Overlap coefficient, an underutilized measure, especially in the context of social networks, is highlighted and shown to achieve a considerably superior efficiency compared to other projection functions. Moreover, we apply concepts to this problem that have not been examined in prior literature, the MinLA problem and the DeGroot model with the presence of stubborn nodes, and show that these techniques are perfectly suitable for the analysis of our dataset. Finally, a relational dataset for the political scene in Greece is assembled, which we believe will inspire other scientists working in the field.

There are several lines of research arising from this work which should be pursued. A natural extension of this work is the political affinity identification of news media and news sources for the purpose of deriving possible political bias present. Previous research demonstrates that the assumptions that our approach is relied upon are well founded in a news media framework. For example, in [32] it is stated that readers have an economically significant preference for like-minded news. This is consistent with our assertions, therefore, a reasonable indication that our methods are suitable for the classification for news sources on Twitter as well. It is, furthermore, worth investigating the scalability of these procedures when applied to the political scenes of other countries.

## Acknowledgements

This work was supported by the project "Assessment of News Reliability in Social Networks of Influence" (Grant no. MIS 5006337) that has been co-financed by the Operational Program "Human Resources Development, Education and Lifelong Learning" and is co-financed by the European Union (European Social Fund) and Greek National funds.

## References

[1] M. Cha, H. Haddadi, F. Benevenuto, and P. K. Gummadi, "Measuring user influence in twitter: The million follower fallacy," in *Proceedings of the 4th International AAAI Conference on Weblogs and Social Media*, ser. ICWSM '10, 2010, pp. 10–17.

[2] H. Kwak, C. Lee, H. Park, and S. Moon, "What is twitter, a social network or a news media?" in *Proceedings of the 19th International Conference on World Wide Web*, ser. WWW '10. New York, NY, USA: ACM, 2010, pp. 591–600.

[3] J. Bollen, H. Mao, and X. Zeng, "Twitter mood predicts the stock market," *Journal of Computational Science*, vol. 2, no. 1, pp. 1 – 8, 2011.

[4] M. Mathioudakis and N. Koudas, "Twittermonitor: Trend detection over the twitter stream," in *Proceedings of the 2010 ACM SIGMOD International Conference on Management of Data*, ser. SIGMOD '10. New York, NY, USA: ACM, 2010, pp. 1155–1158.

[5] Z. Cheng, J. Caverlee, and K. Lee, "You are where you tweet: A content-based approach to geo-locating twitter users," in *Proceedings of the 19th ACM International Conference on Information and Knowledge Management*, ser. CIKM '10. New York, NY, USA: ACM, 2010, pp. 759–768.

[6] R. K. Garrett, "Politically motivated reinforcement seeking: Reframing the selective exposure debate," *Journal of Communication*, vol. 59, no. 4, pp. 676–699.

[7] P. Barberá, "Birds of the same feather tweet together: Bayesian ideal point estimation using twitter data," *Political Analysis*, vol. 23, no. 1, p. 7691, 2015.

[8] A. Tumasjan, T. Sprenger, P. Sandner, and I. Welpe, "Predicting elections with twitter: What 140 characters reveal about political sentiment," in *Proceedings of the 4th International AAAI Conference on Weblogs and Social Media*, ser. ICWSM '10, 2010, pp. 178–185.

[9] M. Pennacchiotti and A.-M. Popescu, "A machine learning approach to twitter user classification," in *Proceedings of the 5th International AAAI Conference on Weblogs and Social Media*, ser. ICWSM '11, 2011, pp. 281–288.

[10] D. Maynard and A. Funk, "Automatic detection of political opinions in tweets," in *The Semantic Web: ESWC 2011 Workshops*, R. García-Castro, D. Fensel, and G. Antoniou, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 88–99.

[11] P. Verweij, "Twitter links between politicians and journalists," *Journalism Practice*, vol. 6, no. 5-6, pp. 680–691, 2012.

[12] A. Boutet, H. Kim, and E. Yoneki, "What's in twitter, i know what parties are popular and who you are supporting now!" *Social Network Analysis and Mining*, vol. 3, no. 4, pp. 1379–1391, Dec 2013.

[13] J. Golbeck and D. Hansen, "A method for computing political preference among Twitter followers," *Social Networks*, vol. 36, pp. 177 – 184, 2014, special Issue on Political Networks.

[14] T. Zhou, J. Ren, M. c. v. Medo, and Y.-C. Zhang, "Bipartite network projection and personal recommendation," *Physical Review E*, vol. 76, p. 046115, Oct 2007.

[15] G. Jeh and J. Widom, "Simrank: A measure of structural-context similarity," in *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '02. New York, NY, USA: ACM, 2002, pp. 538–543.

[16] S. P. Borgatti and D. S. Halgin, "Analyzing affiliation networks," in *The Sage handbook of social network analysis*, J. Scott and P. Carrington, Eds. London, UK: SAGE Publications Ltd, 2014, vol. 1, ch. 28, pp. 417–433.

[17] M. Vijaymeena and K. Kavitha, "A survey on similarity measures in text mining," *Machine Learning and Applications: An International Journal (MLAIJ)*, vol. 3, no. 1, pp. 19–28, 2016.

[18] S. Fortunato, "Community detection in graphs," *Physics Reports*, vol. 486, no. 3, pp. 75 – 174, 2010.

[19] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2008, no. 10, p. P10008, 2008.

[20] A. Lancichinetti and S. Fortunato, "Community detection algorithms: A comparative analysis," *Physical Review E*, vol. 80, p. 056117, Nov 2009.

[21] T. Alzahrani and K. J. Horadam, *Community Detection in Bipartite Networks: Algorithms and Case studies*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2016, pp. 25–50.

[22] M. Bastian, S. Heymann, and M. Jacomy, "Gephi: An open source software for exploring and manipulating networks," in *Proceedings of the 3rd International AAAI Conference on Weblogs and Social Media*, ser. ICWSM '09, 2009, pp. 361–362.

[23] A. Noack, "Modularity clustering is force-directed layout," *Physical Review E*, vol. 79, p. 026102, Feb 2009.

[24] I. Safro, D. Ron, and A. Brandt, "Graph minimum linear arrangement by multilevel weighted edge contractions," *Journal of Algorithms*, vol. 60, no. 1, pp. 24 – 41, 2006.

[25] U. Feige and J. R. Lee, "An improved approximation ratio for the minimum linear arrangement problem," *Information Processing Letters*, vol. 101, no. 1, pp. 26 – 29, 2007.

[26] C. Ambuhl, M. Mastrolilli, and O. Svensson, "Inapproximability results for sparsest cut, optimal linear arrangement, and precedence constrained scheduling," in *48th Annual IEEE Symposium on Foundations of Computer Science (FOCS'07)*, Oct 2007, pp. 329–337.

[27] P. Raghavendra, D. Steurer, and M. Tulsiani, "Reductions between expansion problems," in *IEEE 27th Conference on Computational Complexity*, June 2012, pp. 64–73.

[28] P. McCullagh, "Analysis of ordinal categorical data," *Technometrics*, vol. 27, no. 3, pp. 317–318, 1985.

[29] M. H. Degroot, "Reaching a consensus," *Journal of the American Statistical Association*, vol. 69, no. 345, pp. 118–121, 1974. [Online]. Available: https://amstat.tandfonline.com/doi/abs/10.1080/01621459.1974.10480137

[30] J. Ghaderi and R. Srikant, "Opinion dynamics in social networks: A local interaction game with stubborn agents," in *American Control Conference (ACC)*, June 2013, pp. 1982–1987.

[31] M. Kwiatkowska, G. Norman, and D. Parker, "PRISM 4.0: Verification of probabilistic real-time systems," in *Computer Aided Verification*, ser. LNCS, G. Gopalakrishnan and S. Qadeer, Eds., vol. 6806. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 585–591.

[32] M. Gentzkow and J. M. Shapiro, "What drives media slant? evidence from u.s. daily newspapers," *Econometrica*, vol. 78, no. 1, pp. 35–71.