Spectral Evolution of Twitter Mention Networks

Miguel Romero(✉), Camilo Rocha, and Jorge Finke

Pontificia Universidad Javeriana Cali, Cali, Colombia
{miguel.romero,camilo.rocha,jfinke}@javerianacali.edu.co

Abstract. This papers applies the spectral evolution model presented in [5] to networks of mentions between Twitter users who identified messages with the most popular political hashtags in Colombia (during the period which concludes the disarmament of the Revolutionary Armed Forces of Colombia). The model characterizes the dynamics of each mention network (i.e., how new edges are established) in terms of the eigen decomposition of its adjacency matrix. It assumes that as new edges are established the eigenvalues change, while the eigenvectors remain constant. The goal of our work is to evaluate various link prediction methods that underlie the spectral evolution model. In particular, we consider prediction methods based on graph kernels and a learning algorithm that tries to estimate the trajectories of the spectrum. Our results show that the learning algorithm tends to outperform the kernel methods at predicting the formation of new edges.

Keywords: Spectral evolution model · Twitter mention networks · Eigen decomposition · Graph kernels

1 Introduction

Social networks have become increasingly relevant for understanding the political issues of a country. On such platforms, users share perceptions and opinions on government and public affairs, creating political conversations that often unveil specific patterns of interaction (e.g., the degree of polarization on a current issue). While some studies focus on identifying which profiles play a key role in shaping user-user interactions [9,10], others studies focus on how the user terms and conditions of social networks influence broad political decisions [3,6].

Not surprising, analyzing the patterns that arise from online conversations on social networks has received wide attention [7,8]. Understanding the broad dynamics of user interactions is an important step to evaluate both the formation and political ramifications of stationary patterns. More specifically, characterizing the evolution of user interactions requires the development of models that predict how new edges are established. For example, predicting the formation of new edges is useful to identify whether an influential user retains her status over time or whether a political polarization reflects a dynamic process or a stationary state [2].

H. Cherifi et al. (Eds.): COMPLEX NETWORKS 2019, SCI 881, pp. 532–542, 2020.
https://doi.org/10.1007/978-3-030-36687-2_44

This paper uses the spectral evolution model presented in [5] to capture the dynamics of user interactions and evaluate which link prediction method best estimates the formation of new edges over time. The spectral evolution model considers that the growth of a network can be captured by its eigen decomposition, under the assumption that its eigenvectors remain constant. If this condition is satisfied, the estimation of the formation of new edges can be masked as a transformation of the spectrum through the application of real functions (using graph kernels) or through extrapolation methods (using learning algorithms that estimate the spectrum trajectories) [4].

The main contribution of this paper is to apply the spectral evolution model to networks of mentions between Twitter users who identified messages with the most popular political hashtags H. Vertices represent users and there exists an edge between two users if a user mentions the another user using a hashtag $h \in H$. We select the most popular hashtags related to political affairs in Colombia between August 2017 and August 2018, the period which concludes the disarmament of the Revolutionary Armed Forces of Colombia (Farc) and marks the end of the armed conflict. Different prediction methods are compared to identify which prediction method best describes the evolution of each mention network.

The remainder of the paper is organized as follows. Section 2 describes the networks used for our analysis. Section 3 presents the spectral evolution model and verifies that the model can be applied to the mention networks. Section 3 also overviews the different link prediction methods that underlie the model. Section 4 presents the results of applying the spectral evolution model with various link prediction methods. Section 5 draws some conclusions and future research directions.

2 Data Description

The dataset consists of 31 mention networks between Twitter users who defined their profile location as Colombia. These networks capture conversations around a set of hashtags H related to popular political topics between August 2017 and August 2018. Users are represented by the set of vertices V. The set of edges is denoted by E; there exists an edge $\{i, j\} \in V \times V$ between users i and j, if user i identifies a message with a political hashtag in H (e.g., #safeelections) and mentions user j (via @username). The mention network $G = (V, E)$ is represented as a weighted multi-graph without self-loops, which means that it is possible to have multiple edges between two users. Our analysis is based on the largest connected component of G, denoted by $G_c = (V_c, E_c)$.

A network is built for each hashtag $h \in H$. Table 1 shows a description of the hashtags and the resulting networks, including the number of vertices and edges ($|V|$ and $|E|$) for the whole network G, the number of vertices and edges ($|V_c|$ and $|E_c|$) for its largest component G_c, the community modularity (Q) of G_c, and the number of communities (m) of G_c.

Table 1. Mention networks with political hashtags. English translations for some popular political hashtags appear in parenthesis.

	Set of hashtags H	G		G_c			
		$\|V\|$	$\|E\|$	$\|V_c\|$	$\|E_c\|$	Q	m
0	abortolegalya (legal abortion now)	2235	2202	1282	1538	0.89	30
1	alianzasporlaseguridad (security alliance)	176	1074	150	351	0.34	7
2	asiconstruimospaz (how we build peace)	2514	14055	2405	6950	0.56	16
3	colombialibredefracking (ban fracking)	1606	3483	1476	3127	0.62	19
4	colombialibredeminas (ban mining)	707	2685	655	1421	0.51	14
5	dialogosmetropolitanos (city dialogues)	959	18340	932	4134	0.34	10
6	edutransforma (education transforms)	166	1296	161	404	0.40	9
7	eleccionesseguras (safe elections)	3035	17922	2634	7969	0.51	20
8	elquedigauribe (whoever Uribe says)	2375	6933	2052	5272	0.65	20
9	frutosdelapaz (fruits of peace)	1671	6960	1479	3468	0.58	18
10	garantiasparatodos (assurances for all)	388	814	340	563	0.55	10
11	generosinideologia (no gender ideology)	639	914	615	805	0.63	12
12	hidroituangoescololombia	1028	3362	883	2252	0.68	15
13	horajudicialur (judicial hour)	2250	23647	2187	6756	0.42	14
14	lafauriecontralor (comptroller Lafaurie)	2154	7082	1999	5309	0.59	14
15	lanochesantrich	1518	6946	1444	3567	0.45	13
16	lapazavanza (peace advances)	2949	8288	2775	6569	0.70	18
17	libertadreligiosa (religious liberty)	1584	13443	1395	6856	0.38	15
18	manifestacionpacifica	211	274	112	151	0.69	9
19	plandemocracia2018 (democracy plan)	3090	20955	2962	7996	0.58	22
20	plenariacm (plenary)	1504	19866	1460	4782	0.41	15
21	proyectoituango	1214	3086	1186	1891	0.53	44
22	reformapolitica (political reform)	2714	8385	2608	5928	0.66	18
23	rendiciondecuentas (accountability)	5103	25479	4401	10308	0.84	33
24	rendiciondecuentas2017	1711	12441	998	2933	0.51	16
25	resocializaciondigna	503	4054	496	1171	0.46	8
26	salariominimo (minimum wage)	2494	7041	2079	5016	0.71	22
27	semanaporlapaz (week of peace)	1988	8103	1732	4860	0.69	25
28	serlidersocialnoesdelito	530	861	439	697	0.67	15
29	vocesdelareconciliacion (reconciliation)	161	1500	158	405	0.34	7
30	votacionesseguras (safe voting)	2748	13307	2439	5338	0.66	24

The modularity and number of communities shown in Table 1 are computed with the *multilevel community detection algorithm* [1]. Note that $Q > 0.3$ for all networks in the dataset, i.e., community structure can be observed for all mention networks.

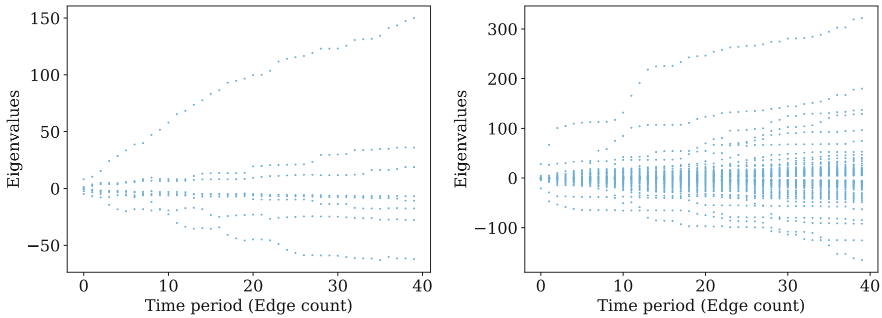
3 Spectral Evolution Model

Let $\mathbf{A}$ denote the adjacency matrix of G_c. Furthermore, let $\mathbf{A} = \mathbf{U}\,\mathbf{\Lambda}\,\mathbf{U}^T$ denote the eigen decomposition of $\mathbf{A}$, where $\mathbf{\Lambda}$ represents the spectrum of G_c. The *spectral evolution model* characterizes the dynamics of G_c (i.e., how new edges are created over time) in terms of the evolution of the spectrum of the network, assuming that its eigenvectors in $\mathbf{U}$ remain unchanged [4,5]. In other words, assume that the dynamics of the network may only involve small changes in behavior of the eigenvectors.

3.1 Spectral Evolution Model Verification

To apply the spectral evolution model, we need to verify the assumption on the evolution of the spectrum and eigenvectors. Every network G_c has a timestamp associated to each edge, representing the time at which the edge was created.

Spectral Evolution. For a given network, the set of edges is split into 40 bins based on their time stamps. Figure 1 illustrates the top 8% of the largest eigenvalues (by absolute value) for two mention networks, namely, #educationtransforms and #howwebuildpeace. For both cases, the eigenvalues grow irregularly, that is, some eigenvalues growth at a higher rate than others. Most of the networks in the dataset show this irregular behavior in spectrum evolution.



(a) #EduTransforma

(b) #AsiConstruimosPaz

Fig. 1. Spectral evolution for mention networks #educationtransforms (left) and #howwebuildpeace (right).

Eigenvector Evolution. At time t, consider the adjacency matrix $\mathbf{A}_{(t)}$, with $1 \leq t \leq T$. The eigenvectors corresponding to the top 8% of the largest eigenvalues (by absolute value) at time t are compared to the eigenvectors at time $T = 40$. In particular, the cosine distance is used as a similarity measure to compare the eigenvectors $\mathbf{U}_{(T)i}$ and $\mathbf{U}_{(t)i}$, for each latent dimension i.

Figure 2 shows that some eigenvectors have a similarity close to one during the entire evolution of the network. These eigenvectors correspond to the eigenvectors associated to the largest eigenvalues. Note also that at some time instants the similarity for some eigenvectors drops to zero, which can be explained because eigenvectors swap locations during eigen decomposition. To identify such changes we verify the stability of the largest eigenvectors.

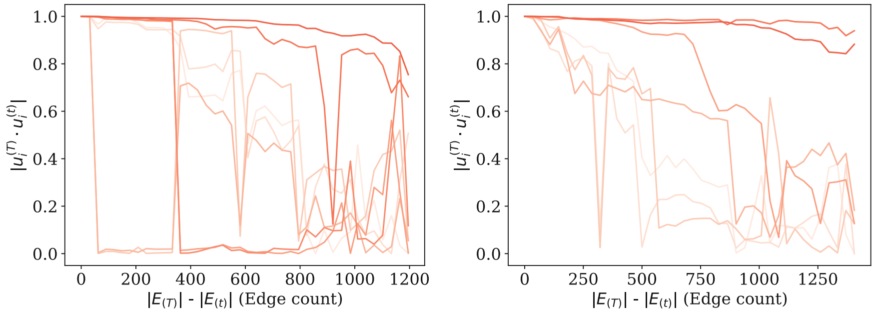
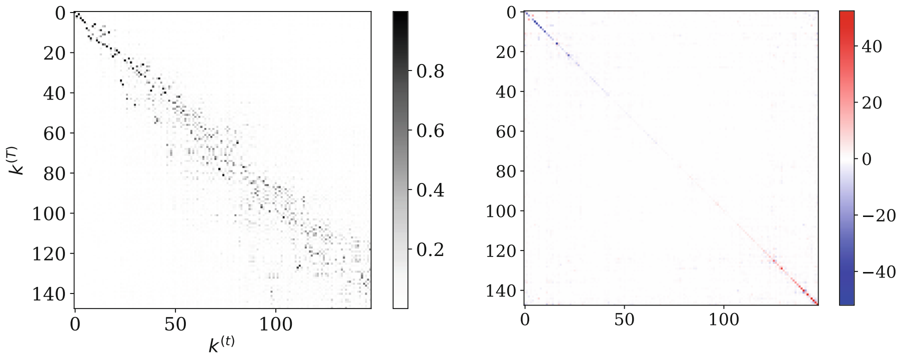


Fig. 2. Eigenvector evolution for mention networks #educationtransforms (left) and #howwebuildpeace (right).

Eigenvector Stability. For a given network G_c, Let t_a and t_b be the times when 75% and 100% of all edges have been created. The eigen decomposition of the adjacency matrices are given by $\mathbf{A}_a = \mathbf{U}_a \, \mathbf{\Lambda}_a \, \mathbf{U}_a^T$ and $\mathbf{A}_b = \mathbf{U}_b \, \mathbf{\Lambda}_b \, \mathbf{U}_b^T$. Similarity values are computed for every pairs of eigenvectors (i, j) using:

$$\text{sim}_{ij}(t_a, t_b) = |\mathbf{U}_{(a)i}^T \cdot \mathbf{U}_{(b)j}|.$$

The resulting values are plotted as a heatmap, where white cells represent a value of zero and black cells a value of one. The more the heatmap approximates a diagonal matrix, the fewer eigenvector permutations there are, i.e., the eigenvectors are preserved over time. Figure 3a shows sub-squares with intermediate values (between zero and one) for the #democracyplan2018 network. These sub-squares result from an exchange in the location of eigenvectors that have eigenvalues that are close in magnitude.



(a) Eigenvector stability (b) Spectral diagonality test

Fig. 3. Eigenvector stability and spectral diagonality test for the #democracyplan2018 network.

Spectral Diagonality Test. As for the eigenvector stability test, consider the eigen decomposition of the adjacency matrix of G_c at time t_a, $\mathbf{A}_a = \mathbf{U}_a \mathbf{\Lambda}_a \mathbf{U}_a^T$. At time $t_b > t_a$ the adjacency matrix is expected to become $\mathbf{A}_b = \mathbf{U}_a(\mathbf{\Lambda}_a + \mathbf{\Delta})\mathbf{U}_a^T$, where $\mathbf{\Delta}$ is a diagonal matrix and indicates whether the growth of the network is spectral.

Using least-squares, the matrix $\mathbf{\Delta}$ can be derived as $\mathbf{\Delta} = \mathbf{U}_a(\mathbf{A}_b - \mathbf{A}_a)\mathbf{U}_a^T$. If $\mathbf{\Delta}$ is diagonal, then the growth between t_a and t_b is spectral. We find that the matrix $\mathbf{\Delta}$ is almost diagonal for all mention networks. Figure 3b, for example, shows the diagonality test for the #democracyplan2018 network.

3.2 Growth Models

Previous sections have verified that the assumptions underlying the spectral evolution model seem to hold to some extent. Broad speaking, eigenvalues grow while eigenvectors remain fairly constant over time.

Next, we consider network growth as a spectral transformation, i.e., in terms of the eigen decomposition of the adjacency matrix. Let $K(\mathbf{A})$ be a kernel of an adjacency matrix $\mathbf{A}$, whose eigen decomposition is $\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$. Graph kernels assume that there exists a real function $f(\lambda)$ that describes the growth of the spectrum. In particular, $K(\mathbf{A})$ can be written as $K(\mathbf{A}) = \mathbf{U}F(\mathbf{\Lambda})\mathbf{U}^T$, for some function $F(\mathbf{\Lambda})$ that applies a real function $f(\lambda)$ to the eigenvalues of $\mathbf{A}$. In particular, we use the triangle closing kernel, the exponential kernel, and the Neumann growth kernel.

Triangle Closing Kernel. The triangle closing kernel is expressed as $\mathbf{A}^2 = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T\mathbf{U}\mathbf{\Lambda}\mathbf{U}^T = \mathbf{U}\mathbf{\Lambda}^2\mathbf{U}^T$, since $\mathbf{U}^T\mathbf{U} = \mathbf{I}$. This spectral transformation replaces the eigenvalues of $\mathbf{A}$ by their squared values. The real function associated to the triangle closing kernel is $f(\lambda) = \lambda^2$.

Exponential Kernel. The exponential of the adjacency matrix $\mathbf{A}$ is called the exponential kernel. This kernel denotes the sum of every path between two vertices weighted by the inverse factorial of its length. It is expressed as

$$\exp(\alpha\mathbf{A}) = \sum_{k=0}^{\infty} \alpha^k \frac{1}{k!} \mathbf{A}^k,$$

where α is a constant used to balance the weight of short and long paths. The real function associated to the exponential kernel is $f(\lambda) = e^{\alpha\lambda}$.

Neumann Kernel. The Neumann kernel is expressed as

$$(\mathbf{I} - \alpha\mathbf{A})^{-1} = \sum_{k=0}^{\infty} \alpha^k \mathbf{A}^k,$$

where $\alpha^{-1} > |\lambda_1|$ and λ_1 is the largest eigenvalue of $\mathbf{A}$. Its real function is given by $f(\lambda) = 1/(1 - \alpha\lambda)$.

Spectral Extrapolation. As noted above, graph kernels assume that there exists a real function $f(\lambda)$ that describes the growth of the spectrum. However, when the evolution of the spectrum is irregular, as in Fig. 1a, it is not possible to find a simple function that describe network growth. The spectral extrapolation method is a generalization of the graph kernels, which extrapolates each eigenvalue under the assumption that the network follows the spectral evolution model [4].

More specifically, given a network with a timestamped set of edges, the set is split into three subsets named training, target and test sets. Consider two time instants t_a and t_b. Let A_a represent the adjacency matrix of the network at time t_a and $A_a + A_b$ the adjacency matrix at time t_b. The eigen decompositions of the network at the two time instances are given by $\mathbf{A}_a = \mathbf{U}_a \, \mathbf{\Lambda}_a \, \mathbf{U}_a^T$ and $\mathbf{A}_a + \mathbf{A}_b = \mathbf{U}_b \, \mathbf{\Lambda}_b \, \mathbf{U}_b^T$.

Next, let $(\lambda_b)_j$ be the j-eigenvalue at time t_b. Its previous value at time t_a is estimated as a diagonalization of $\mathbf{A}_a$ by $\mathbf{U}_b$ as follows:

$$(\hat{\lambda}_a)_j = \left(\sum_i (\mathbf{U}_a)_i^T (\mathbf{U}_a)_j \right)^{-1} \sum_i (\mathbf{U}_a)_i^T (\mathbf{U}_a)_j (\lambda_a)_i,$$

where $(\mathbf{U}_a)_i$ and $(\lambda_a)_i$ are the eigenvectors and eigenvalues of $\mathbf{A}$, respectively. A linear extrapolation is now performed to predict the eigenvalues $(\hat{\lambda}_c)_i$ at a future time t_c,

$$(\hat{\lambda}_c)_j = 2(\lambda_b)_j - (\hat{\lambda}_a)_j.$$

The predicted matrix $\hat{\mathbf{\Lambda}}_c$ is used to compute the predicted edge weights $\hat{\mathbf{A}}_c = \mathbf{U}_b \, \hat{\mathbf{\Lambda}}_c \, \mathbf{U}_b^T$.

4 Case Study: Twitter Conversations

This section presents the results of applying the proposed kernels (namely, triangle closing, exponential, and Neumann kernels) and the extrapolation method to predict the creation of new edges across the mention networks described in Sect. 2. Curve-fitting methods are applied to find the parameters α of the exponential and Neumann kernels.

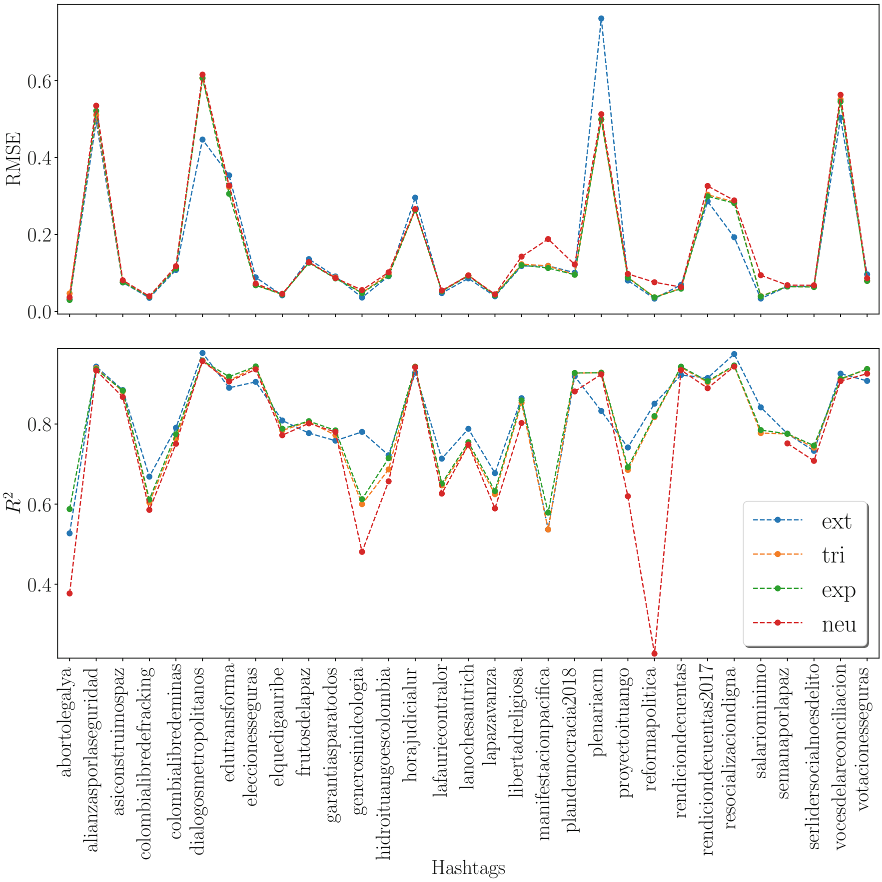


Fig. 4. Performance of the prediction of the methods is evaluated based on two metrics, RMSE and R^2.

To evaluate the performance of the methods we compute the metrics of the root mean square error (RMSE) and R^2.

Figure 4 summarizes the result of RMSE and R^2 metrics. Note that the performance of the models appear to be very similar for most mention networks. In Sect. 3, we verify that the growth of the eigenvalues for most networks is irregular. It is therefore to some extent expected that the extrapolation method outperform the graph kernels. Next, we borrow the structural similarity index method (SSIM) from the filed of image processing to measure the similarity between the actual and the estimated adjacency matrices. (SSIM is widely applied in the field of image processing to compare the similarity between two images based on the idea that pixels have strong inter-dependencies when they are spatially close [11].) Unlike other techniques, such as RMSE, SSIM relies on the estimation of point-to-point absolute errors.

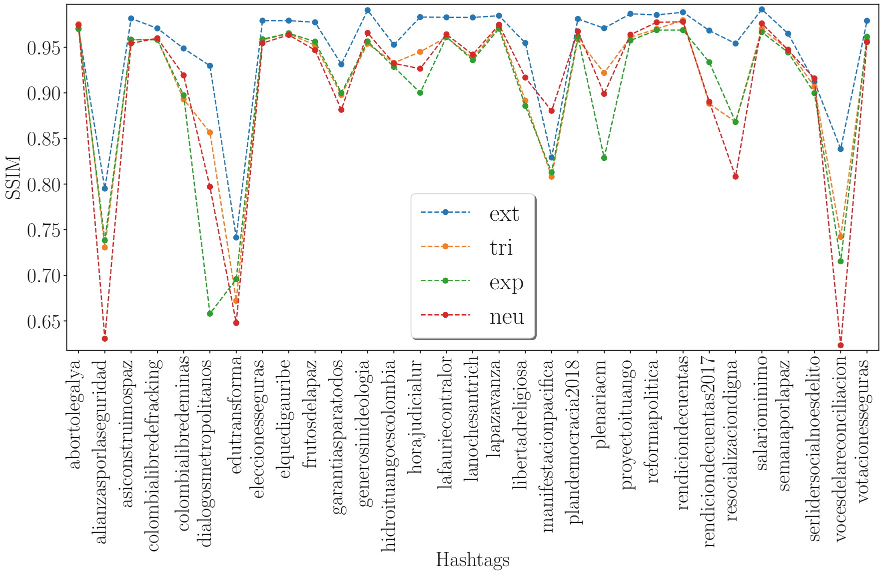


Fig. 5. Performance of the prediction of the methods is evaluated based on SSIM method.

The results are shown in Table 2 and Fig. 5. Figure 5 summarizes the performance for all methods using SSIM. In general, the extrapolation method tends to outperform the other methods. Specifically, for 28 out of 31 networks (91% of the total), the extrapolation method provides a distinct, if sometimes slight, improvement. The Neumann kernel and the triangle closing combined provide better estimates only for 3 networks.

Whenever the spectral extrapolation method outperforms the graph kernels, better prediction seem to be explained by the method being able to consider the irregular evolution of the eigenvalues. In general, note that the networks considered are large enough so that only a small number of eigenvalues and eigenvectors can be computed.

Table 2. Spectral evaluation model performance analysis with SSIM.

	Hashtag	extrapol	$\mathbf{A}^2$	$\exp(\alpha\mathbf{A})$	$(\mathbf{I}-\alpha\mathbf{A})^{-1}$	Best kernel or method
0	abortolegalya	0.97	0.98	0.97	0.97	$\mathbf{A}^2$
1	alianzasporlaseguridad	0.80	0.73	0.74	0.63	extrapol.
2	asiconstruimospaz	0.98	0.96	0.96	0.95	extrapol.
3	colombialibredefracking	0.97	0.96	0.96	0.96	extrapol.
4	colombialibredeminas	0.95	0.89	0.90	0.92	extrapol.
5	dialogosmetropolitanos	0.93	0.86	0.66	0.80	extrapol.
6	edutransforma	0.74	0.67	0.70	0.65	extrapol.
7	eleccionesseguras	0.98	0.96	0.96	0.95	extrapol.
8	elquedigauribe	0.98	0.96	0.97	0.96	extrapol.
9	frutosdelapaz	0.98	0.95	0.96	0.95	extrapol.
10	garantiasparatodos	0.93	0.90	0.90	0.88	extrapol.
11	generosinideologia	0.99	0.95	0.96	0.97	extrapol.
12	hidroituangoescolombia	0.95	0.93	0.93	0.93	extrapol.
13	horajudicialur	0.98	0.94	0.90	0.93	extrapol.
14	lafauriecontralor	0.98	0.96	0.96	0.96	extrapol.
15	lanochesantrich	0.98	0.94	0.94	0.94	extrapol.
16	lapazavanza	0.98	0.97	0.97	0.97	extrapol.
17	libertadreligiosa	0.95	0.89	0.89	0.92	extrapol
18	manifestacionpacifica	0.83	0.81	0.81	0.88	$(\mathbf{I}-\alpha\mathbf{A})^{-1}$
19	plandemocracia2018	0.98	0.96	0.96	0.97	extrapol.
20	plenariacm	0.97	0.92	0.83	0.90	extrapol.
21	proyectoituango	0.99	0.96	0.96	0.96	extrapol.
22	reformapolitica	0.99	0.97	0.97	0.98	extrapol.
23	rendiciondecuentas	0.99	0.98	0.97	0.98	extrapol.
24	rendiciondecuentas2017	0.97	0.89	0.93	0.89	extrapol.
25	resocializaciondigna	0.95	0.87	0.87	0.81	extrapol.
26	salariominimo	0.99	0.97	0.97	0.98	extrapol.
27	semanaporlapaz	0.96	0.95	0.94	0.95	extrapol.
28	serlidersocialnoesdelito	0.91	0.91	0.90	0.92	$(\mathbf{I}-\alpha\mathbf{A})^{-1}$
29	vocesdelareconciliacion	0.84	0.74	0.72	0.62	extrapol.
30	votacionesseguras	0.98	0.96	0.96	0.96	extrapol.

5 Conclusions

This paper applies the spectral evolution model to 31 Twitter mention networks. This model characterizes the evolution of each network in terms of the eigen decomposition of its adjacency matrix. It has been verified that Twitter mention networks follow the spectral evolution model. For most networks, the eigenvectors remain approximately constant, while the spectra of the mention networks grow irregularly. Their evolution can be predicted with the help different growth models. Our results shows that the extrapolation method outperforms the kernel methods mainly due to the irregular evolution of the spectra. Developing more refined models that use learning to predict the evolution of the spectra of graphs remains an important direction for future research.

Acknowledgements. This work was funded by the OMICAS program: Optimización Multiescala In-silico de Cultivos Agrícolas Sostenibles (Infraestructura y Validación en Arroz y Caña de Azúcar), sponsored within the Colombian Scientific Ecosystem by the World Bank, Colciencias, Icetex, the Colombian Ministry of Education, and the Colombian Ministry of Industry and Turism, under GRANT ID: FP44842-217-2018.

References

1. Blondel, V.D., Guillaume, J.-L., Lambiotte, R., Lefebvre, E.: Fast unfolding of communities in large networks. J. Stat. Mech: Theory Exp. **2008**(10), P10008 (2008)
2. DiMaggio, P., Evans, J., Bryson, B.: Have american's social attitudes become more polarized? Am. J. Sociol. **102**(3), 690–755 (1996)
3. Gustafsson, N.: The subtle nature of Facebook politics: Swedish social network site users and political participation. New Media Soc. **14**(7), 1111–1127 (2012)
4. Kunegis, J., Fay, D., Bauckhage, C.: Network growth and the spectral evolution model. In: Proceedings of the 19th ACM International Conference on Information and Knowledge Management, Toronto, ON, Canada, p. 739. ACM Press (2010)
5. Kunegis, J., Fay, D., Bauckhage, C.: Spectral evolution in dynamic networks. Knowl. Inf. Syst. **37**(1), 1–36 (2013)
6. Loader, B.D., Mercea, D.: Networking democracy?: Social media innovations and participatory politics. Inf. Commun. Soc. **14**(6), 757–769 (2011)
7. Mcclurg, S.D.: Social networks and political participation: the role of social interaction in explaining political participation. Polit. Res. Q. **56**(4), 449–464 (2003)
8. McPherson, M., Smith-Lovin, L., Cook, J.M.: Birds of a feather: homophily in social networks. Ann. Rev. Sociol. **27**(1), 415–444 (2001)
9. Noveck, B.S.: Five hacks for digital democracy. Nature **544**(7650), 287–289 (2017)
10. Persily, N.: Can democracy survive the Internet? J. Democracy **28**(2), 63–76 (2017)
11. Wang, Z., Bovik, A., Sheikh, H., Simoncelli, E.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**(4), 600–612 (2004)