INTER-FACULTY MASTER PROGRAM on
COMPLEX SYSTEMS and NETWORKS
SCHOOL of MATHEMATICS
SCHOOL of BIOLOGY
SCHOOL of GEOLOGY
SCHOOL of ECONOMICS
ARISTOTLE UNIVERSITY of THESSALONIKI

Information Technologies Institute

Unsupervised keyword extraction using the GoW model and centrality scores

Elissavet Batziou, Ilias Gialampoukidis, Stefanos Vrochidis,
Ioannis Antoniou, Ioannis Kompatsiaris

Thessaloniki, November 2017

Presenter: Elissavet Batziou batziou.el@gmail.com

OUTLINE

- Graph centralities and community detection
- Bag of words and graph of words
- Application to keyword extraction
- Contribution
  - We examined the performance of 17 keyword extraction techniques based on centrality measures and community detection approaches on the graph of words.
  - We also proposed Mapping Entropy Closeness (MEC) centrality measure

GRAPH CENTRALITIES

- **Graph formulation**

Given an undirected network $G(N, L)$ with N nodes and L links, the adjacency matrix $\boldsymbol{A}$ of a network $G(N, L)$ is a square matrix which is defined as follows:

$$A(n_i, n_j) = A_{ij} = \begin{cases} 1, if\ n_i, n_j\ are\ connected \\ 0 \qquad\qquad otherwise \end{cases}$$

In general, we denote by M_{ij} the (i, j) element of a matrix M.

- **Degree centrality**

Degree of a node n_k, $\deg(n_k)$ is the number of edges connected to it. The maximum number of nodes that node n_k can be connected is $N-1$ and the degree centrality (DC) of node n_k is defined as (Freeman, 1979):

$$DC_k = \frac{\deg(n_k)}{N-1}$$

- **Betweenness centrality**

Let n_i, n_j be two nodes and g_{ij} the number of geodesics linking n_i with n_j

Let also $g_{ij}(n_k)$ the number of geodesics linking n_i and n_j that contain n_k

The betweenness centrality of node n_k (Freeman,1977):

$$BC_k = \frac{2\sum_{i<j}^{N} \frac{g_{ij}(n_k)}{g_{ij}}}{N^2 - 3N + 2}$$

- **Closeness centrality**

Let $d(n_i, n_k)$ be the number of edges in the geodesic linking n_i and n_j.

$$Decentrality(farness) = \frac{\sum_{i=1}^{N} d(n_i, n_k)}{N-1}$$

and the closeness centrality CC of the node n_k is defined as:

$$CC_k = \frac{1}{\text{decentrality}} = \frac{N-1}{\sum_{i=1}^{N} d(n_i, n_k)}$$

- **Eigenvector centrality**

Let $x = \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix}$ be a vector where x_k the centrality of node n_k.

The centrality x_k of node n_k depends on the n_k's network neighbors centrality:

$$x_k = \frac{1}{\lambda} \sum_{j=1}^{N} A_{kj}\, x_j \Leftrightarrow \lambda x = Ax$$

- **Page Rank** centrality of node n_k is defined as:

$$PR_k = \frac{1-d}{N} + d \sum_{n_i \in \mathcal{N}(n_k)} \frac{PR_i}{L(n_i)}$$

where d is the damping factor, typically set to 0.085, $L(n_i)$ is the number of links to node n_i and $\mathcal{N}(n_k)$ is the neighborhood of n_k.

- **Mapping Entropy**

The set of nodes connected to node n_k, $\mathcal{N}(n_k)$ has been used to define the mapping entropy (ME) centrality (Nie et al., 2016) as a function of the degree centrality:

$$ME_k = -DC_k \sum_{n_i \in \mathcal{N}(n_k)} \log DC_i$$

- **Mapping Entropy Betweenness**

Mapping Entropy has been extended (Gialampoukidis et al., 2016) by replacing the degree centrality with the Betweenness centrality, as follows:

$$MEB_k = -BC_k \sum_{n_i \in \mathcal{N}(n_k)} \log BC_i$$

- **Mapping Entropy Closeness**

Mapping Entropy Closeness (MEC) centrality of node n_k is an other extension of Mapping Entropy which is proposed in this thesis:

$$MEC_k = -CC_k \sum_{n_i \in \mathcal{N}(n_k)} \log CC_i$$

▪ Coreness

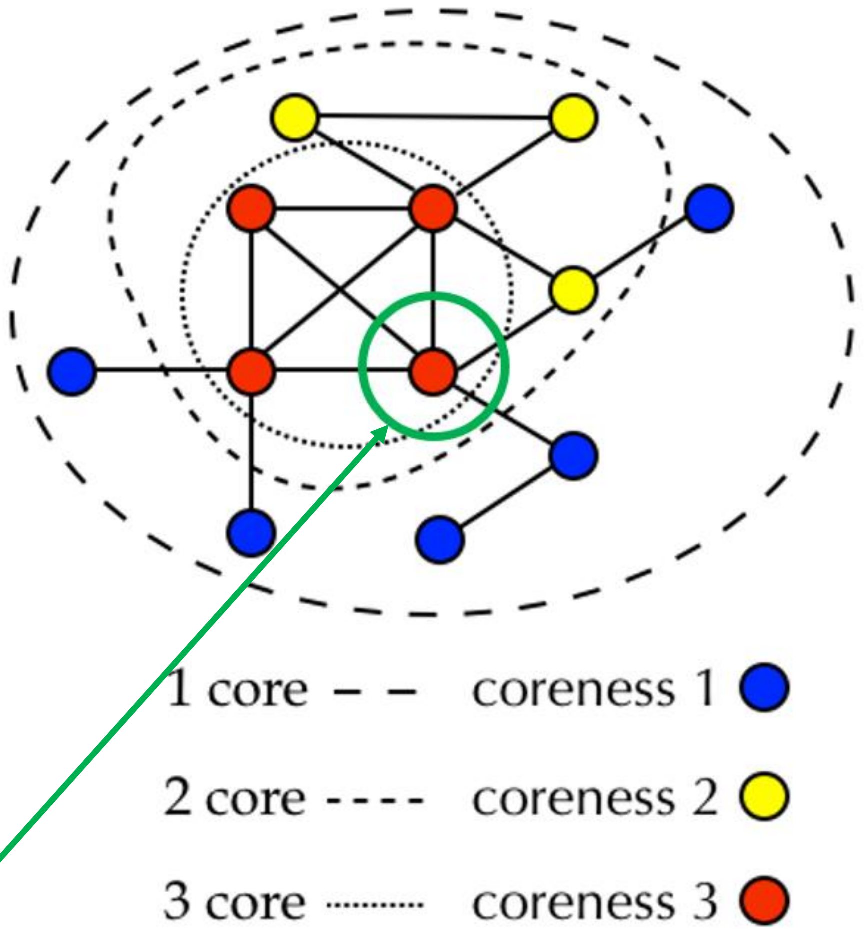
The k-core of a graph G is defined as the maximum subgraph of G in which all nodes have at least degree k.

The coreness of a node of the graph G is k if it belongs to the k-core but not to the $(k+1)$-core.

▪ Eccentricity

The eccentricity of a node k in a graph G is the greatest geodesic distance between the node k and any other node.

The eccentricity can be considered as a centrality measure because the most central node of a graph has the minimum eccentricity.



The selected node (in the green circle) is the node with the minimum eccentricity

- **Clustering Coefficient** (local transitivity)

The local clustering coefficient of a node n_i in a graph G quantifies how close the neighbors of n_i are to being a clique (complete graph).

The local clustering coefficient of a node n_i in an undirected graph is defined as:

$$C_i = \frac{2|\{e_{jk}: n_j, n_k \in N_{i,} e_{jk} \in E\}|}{\deg(n_i)(\deg(n_i) - 1)}$$

where e_{jk} is the link from node n_j to n_k, N_i is the set of neighbours of n_i.

COMMUNITY DETECTION

- **Girvan-Newman algorithm**

GN algorithm is based on the edge betweenness centrality measure.

The edge betweenness determines the edges which are more possible to link different communities.

In order to extract communities, the modularity score is computed, so as to be maximized (Newman and Girvan, 2004):

$$Q = \sum_i (e_{ii} - a_i^2), \qquad a_i = \sum_j e_{ij}$$

where e_{ij} are the elements of a $k \times k$ symmetric matrix and k is the number of communities at which the graph is partitioned.

The elements e_{ij} are defined as the fraction of all edges in the network that link vertices in community i to vertices in community j.

- **Fast Greedy algorithm** (modularity maximization)

All nodes are separate communities and any two communities are merged if the modularity increases.

The algorithm stops when the modularity is not increasing anymore.

The modularity function is defined as (Clauset et al., 2004):

$$Q = \frac{1}{2L} \sum_{i,j} \left[A_{ij} - \frac{\deg(n_i)\deg(n_j)}{2L} \right] \delta(i,j)$$

where L is the number of links in the graph and $\delta(i,j)$ is 1 if $i = j$ and 0 otherwise.

The modularity maximization algorithm of (Clauset et al., 2004) is a faster method to detect communities based on the modularity maximization, compared to the Girvan–Newman community detection algorithm.

- **Louvain method**

The Louvain method (Blondel et al., 2008) is based on the maximization of the modularity Q and involves two phases that are repeated iteratively.

In the first phase, each node forms a community and for each node i the gain of modularity is calculated for removing vertex i from its own community and placing it into the community of each neighbor j of i.

The vertex i is moved to the community for which the gain in modularity becomes maximal.

The first phase is completed when the modularity cannot be further increased.

In the second phase, the detected communities formulate a new network with weights of the links between the new nodes being the sum of weights of the links between nodes in the corresponding two communities.

In this new network, self-loops are allowed, representing links between vertices of the same community.

At the end of the second phase, the first phase is re-applied to the new network, until no more communities are merged and the modularity attains its maximum.

Infomap method

Infomap method (Rosvall and Bergstrom, 2008; Rosvall et al., 2010) minimizes the Shannon information (Cover and Thomas, 2012) required to describe the trajectory of a random walk on the network.

Let ξ be a network partition into m communities

Aim: Codelength $\ell(\xi)$ minimization among all possible partitions ξ of the network:

$$\ell(\xi) = q_{\curvearrowleft}\mathcal{I}(Q) + \sum_{i=1}^{m} p_{i\circlearrowright}\,\mathcal{I}(\mathcal{P}_i)$$

where $q_{\curvearrowleft} = \sum_{i=1}^{m} q_{i\curvearrowleft}$

$q_{i\curvearrowleft}$ the rate at which the random walk enters community-i

Q the probability distribution of $q_{i\curvearrowleft}$

$p_{i\circlearrowright}$ the rate at which the random walk uses community-i

$\mathcal{P}_i$ the probability distribution of $p_{i\circlearrowright}$

- **Label propagation**

The Label Propagation method (Raghavan et al., 2007) initializes every node with a unique label and at each step every node adopts the label that most of its neighbors currently have.

Hence, an iterative process is defined, in which densely connected groups of nodes form a consensus on a label and communities are extracted.

- **Walktrap method**

The Walktrap method (Pons and Latapy, 2005) generates random short walks on the graph by simulating transitions between nodes.

Since short random walks tend to stay within the same community, it is possible to detect communities using such random walks.

BAG AND GRAPH OF WORDS

In the BoW model, a text document is represented as a vector, containing all text's words free from grammar and word order.

Word's multiplicity is the number of occurrences of a word in a document, known also as **term frequency** (tf):

$$tfidf_{ij} = \frac{n_{id}}{n_d} \log \frac{N}{n_i}$$

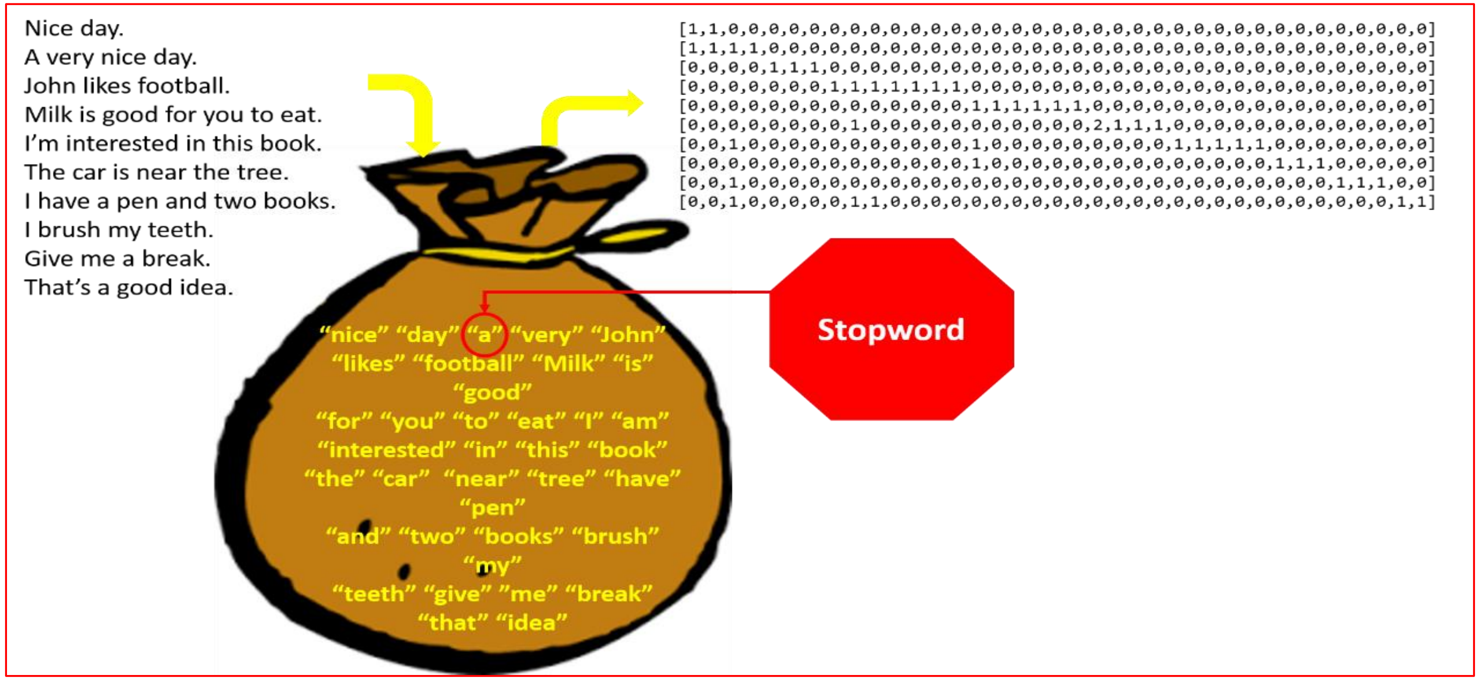
n_{id} = the number of occurrences of word i in document d

n_d = the number of words in document d

n_i = the number of occurrences of word i in the whole database

N = the total number of documents in the database

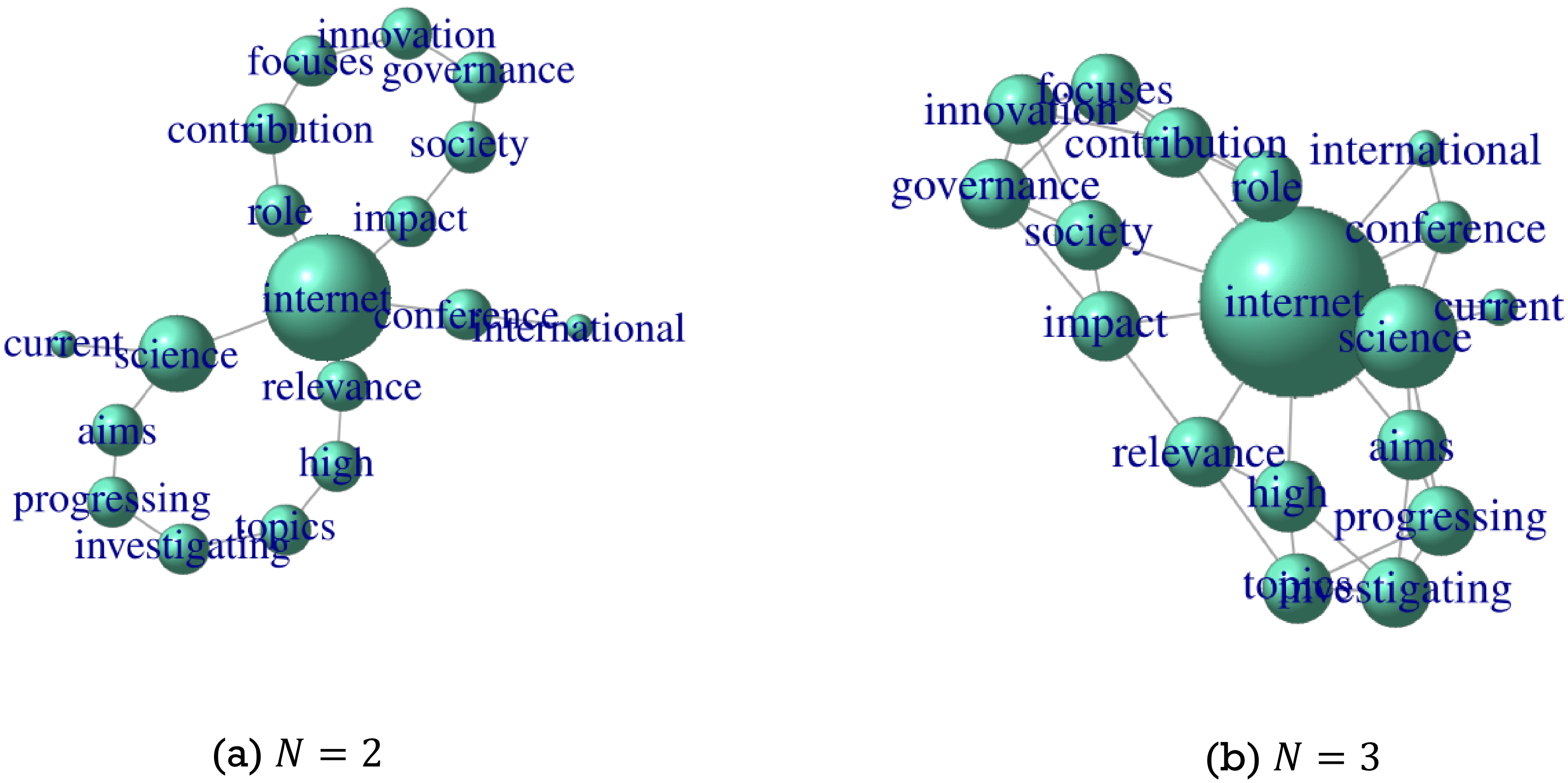
$\frac{n_{id}}{n_d}$ is the term frequency



However, it is possible to consider pairs, triplets or n-tuples of words as "terms", known as word n-grams.

Graph of words (GoW) model (Rousseau and Vazirgiannis, 2013)

Given a window of N successive words in a document, all terms in the window are mutually linked and each edge represents the co-occurrence of a pair of terms.



Graph of Words for $N = 2$ and $N = 3$ on the text "The international conference on Internet Science aims at progressing and investigating on topics of high relevance with Internet's impact on society, governance, and innovation. It focuses on the contribution and role of Internet science on the current…"

APPLICATION TO KEYWORD EXTRACTION

METHODS

- Betweenness centrality
- Closeness centrality
- Degree centrality
- Eigenvector centrality
- PageRank
- Eccentricity
- Coreness
- Transitivity
- Mapping Entropy
- Mapping Entropy Betweenness
- Mapping Entropy Closeness

- Fast greedy (modularity maximization)
- Infomap (codelength minimization)
- Label Propagation
- Louvain (modularity maximization)
- Walktrap (random walks)

- Term-Frequency (TF) scores

EVALUATION MEASURES

Let $\mathcal{C}$ be the collection of documents and we denote by $\mathcal{R}$ the set of retrieved results with respect to the query q. We also denote by $\mathcal{T}$ the set of relevant documents, in terms of the annotation which is provided by the ground truth.

- **Precision**

$$precision = \frac{|relevant\ documents| \cap |retrieved\ documents|}{|retrieved\ documents|} = \frac{|\mathcal{T} \cap \mathcal{R}|}{|\mathcal{R}|}$$

- **Average Precision**

$$AP = \frac{\sum_{n=1}^{R} P@n}{R}$$

where n is the rank of each relevant document and R is the total number of relevant documents.

- $\boldsymbol{P@n}$ is the precision of the top-n retrieved documents

- **Mean Average Precision**

the mean of all Average Precision scores for each query:

$$mAP = \frac{\sum_{q=1}^{Q} AP(q)}{Q}$$

where, $AP(q)$ is the Average Precision for the query q.

- **Jaccard similarity**

The Jaccard index, also known as the Jaccard similarity coefficient, is a statistic used for comparing the similarity of two sample sets and is defined as the size of the intersection divided by the size of the union of the sample sets:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

EXPERIMENTS

DATASETS

FAO sample text

Where to purchase FAO publications locally - Points de vente des publications de la FAO - Puntos de venta de publicaciones de la FAO
· ANGOLA
Empresa Nacional do Disco e de Publicaɳυes, ENDIPU-U.E.E.
Rua Cirilo da Conceiɳγo Silva, N° 7
C.P. N° 1314-C, Luanda
· ARGENTINA
Librerva Agropecuaria
Pasteur 743, 1028 Buenos Aires
Oficina del Libro Internacional
Av. Cσrdoba 1877, 1120 Buenos Aires
E-mail: olilibro@satlink.com
· AUSTRALIA
Hunter Publications
P.O. Box 404, Abbotsford, Vic. 3067
Tel.:(03) 9417 5361
Fax: (03) 914 7154
E-mail: jpdavies@ozemail.com.au
· AUSTRIA
Gerold Buch & Co.
Weihburggasse 26, 1010 Vienna

CiteULike sample text

The study of networks pervades all of science, from neurobiology to statistical physics. The most basic issues are structural: how does one characterize the wiring diagram of a food web or the Internet or the metabolic network of the bacterium Escherichia coli? Are there any unifying principles underlying their topology? From the perspective of nonlinear dynamics, we would also like to understand how an enormous network of interacting dynamical systems -- be they neurons, power stations or lasers -- will behave collectively, given their individual dynamics and coupling architecture. Researchers are only now beginning to unravel the structure and dynamics of complex networks. Networks are on our minds nowadays. Sometimes we fear their power -- and with good reason. On 10 August 1996, a fault in two power lines in Oregon led, through a cascading series of failures, to blackouts in 11 US states and two Canadian provinces, leaving about 7 million customers without power for up to 16 hours1. The Love Bug worm, the worst computer attack to date, spread over the Internet on 4 May 2000 and inflicted billions of dollars of damage worldwide. In our lighter moments we play parlour games about connectivity.

The CiteULike dataset has 183 publications crawled from CiteULike, and keywords assigned by 152 different CiteULike users who saved these publications. The other dataset, FAO780, has 779 FAO publications with Agrovoc terms from official documents of the Food and Agriculture Organization of the United Nations (FAO).

SETTINGS

- remove punctuation
- transform all letters to lowercase
- Numbers are removed
- English stopwords are removed
- we stem each word
- we construct the graph of words, which has as nodes the words of each document

In all datasets, we keep the top-20 keywords for each selected centrality score and for the top-20 most frequent terms (TF scores).

RESULTS

N=2	Citeulike180			Fao780		
Method	Jaccard	Average Precision	P@10	Jaccard	Average Precision	P@10
Betweenness	0.1531 ± 0.0598	0.3795 ± 0.1401	0.3486 ± 0.1398	0.1619 ± 0.0734	0.3459 ± 0.1500	0.3112 ± 0.1473
Closeness	0.1531 ± 0.0622	**0.3890** ± 0.1425	**0.3552** ± 0.1413	0.1656 ± 0.0781	0.3565 ± 0.1547	0.3212 ± 0.1540
Degree	0.1566 ± 0.0611	0.3842 ± 0.1390	0.3492 ± 0.1410	0.1671 ± 0.0777	0.3533 ± 0.1538	0.3208 ± 0.1508
Eigenvector	0.1446 ± 0.0659	0.3606 ± 0.1453	0.3525 ± 0.1421	0.1649 ± 0.0792	0.3526 ± 0.1570	0.3158 ± 0.1549
Page Rank	0.0508 ± 0.0313	0.3831 ± 0.1399	0.3492 ± 0.1410	0.1669 ± 0.0772	0.3488 ± 0.1530	0.3173 ± 0.1503
Mapping Entropy	0.1557 ± 0.0613	0.3821 ± 0.1394	0.3519 ± 0.1406	0.1669 ± 0.0780	0.3515 ± 0.1533	0.3191 ± 0.1502
MEB	**0.1598** ± 0.0625	0.3860 ± 0.1378	0.3530 ± 0.1354	0.0674 ± 0.0451	0.1762 ± 0.1180	0.1469 ± 0.1009
MEC	0.1567 ± 0.0622	0.3839 ± 0.1389	0.3503 ± 0.1402	0.0678 ± 0.0460	0.1753 ± 0.1178	0.1477 ± 0.1009
Coreness	0.1098 ± 0.5110	0.2857 ± 0.1364	0.3508 ± 0.1568	0.0839 ± 0.0487	0.1802 ± 0.0994	0.2855 ± 0.1556
Transitivity	0.0000 ± 0.0000	0.0182 ± 0.0469	0.0164 ± 0.0426	0.0067 ± 0.0154	0.0221 ± 0.0559	0.0171 ± 0.0422
Eccentricity	0.0015 ± 0.0062	0.0026 ± 0.0157	0.0027 ± 0.0163	0.0003 ± 0.0033	0.0004 ± 0.0054	0.0004 ± 0.0062
TF score	0.1613 ± 0.0648	0.3877 ± 0.1421	0.3530 ± 0.1386	**0.1781** ± 0.0843	**0.3725** ± 0.1603	**0.3392** ± 0.1614
Fast greedy	0.0215 ± 0.0164	0.0649 ± 0.0500	0.1656 ± 0.1459	0.0100 ± 0.0116	0.0297 ± 0.0303	0.1163 ± 0.1114
Infomap	0.0402 ± 0.0248	0.1258 ± 0.0762	0.2749 ± 0.1770	0.0205 ± 0.0220	0.0586 ± 0.0581	0.2258 ± 0.1462
Label Prop	0.0158 ± 0.0088	0.0411 ± 0.0203	0.2754 ± 0.1693	0.0074 ± 0.0069	0.0219 ± 0.0153	0.2100 ± 0.1420
Louvain	0.0193 ± 0.0167	0.0600 ± 0.0538	01421 ± 0.1415	0.0107 ± 0.0130	0.0320 ± 0.0359	0.0992 ± 0.1054
Walktrap	0.0332 ± 0.0171	0.0941 ± 0.0459	0.3060 ± 0.1846	0.0176 ± 0.0173	0.0504 ± 0.0412	0.2144 ± 0.1439

N=3	Citeulike180			Fao780		
Method	Jaccard	Average Precision	P@10	Jaccard	Average Precision	P@10
Betweenness	0.1609 ± 0.0633	0.3854 ± 0.1431	0.3519 ± 0.1441	0.1671 ± 0.0748	0.3568 ± 0.1505	0.3213 ± 0.1504
Closeness	$\mathbf{0.1658} \pm 0.0617$	$\mathbf{0.4034} \pm 0.1447$	$\mathbf{0.3776} \pm 0.1490$	0.1731 ± 0.0819	0.3678 ± 0.1560	0.3326 ± 0.1558
Degree	0.1648 ± 0.0621	0.3993 ± 0.1406	0.3661 ± 0.1404	0.1744 ± 0.0806	0.3671 ± 0.1543	0.3304 ± 0.1532
Eigenvector	0.1542 ± 0.0629	0.3791 ± 0.1445	0.3448 ± 0.1428	0.1711 ± 0.0818	0.3662 ± 01589	0.3291 ± 0.1590
Page Rank	0.1645 ± 0.0662	0.3982 ± 0.1401	0.3678 ± 0.1395	0.1740 ± 0.0807	0.3641 ± 0.1542	0.3286 ± 0.1530
Mapping Entropy	0.1644 ± 0.0632	0.3974 ± 0.1404	0.3650 ± 0.1394	0.1746 ± 0.0807	0.3662 ± 0.1544	0.3295 ± 0.1540
MEB	0.1638 ± 0.0619	0.3963 ± 0.1397	0.3661 ± 0.1435	0.1723 ± 0.0776	0.3627 ± 0.1527	0.3293 ± 0.1530
MEC	0.1648 ± 0.0636	0.3886 ± 0.1407	0.3683 ± 0.1402	0.1745 ± 0.0803	0.3671 ± 0.1544	0.3295 ± 0.1527
Coreness	0.1066 ± 0.0481	0.2637 ± 0.1208	0.3694 ± 0.1682	0.075 ± 0.0440	0.1595 ± 0.0848	0.2796 ± 0.1542
Transitivity	0.0015 ± 0.0062	0.0025 ± 0.0161	0.0022 ± 0.0147	0.0001 ± 0.0050	0.0015 ± 0.0130	0.0014 ± 0.0118
Eccentricity	0.0016 ± 0.0067	0.0022 ± 0.0124	0.0033 ± 0.0179	0.0006 ± 0.0045	0.0010 ± 0.0090	0.0006 ± 0.0080
TF score	0.1613 ± 0.0648	0.2637 ± 0.1208	0.3530 ± 0.1386	$\mathbf{0.1781} \pm 0.0843$	$\mathbf{0.3725} \pm 0.1603$	$\mathbf{0.3392} \pm 0.1614$
Fast greedy	0.0196 ± 0.0146	0.0565 ± 0.0399	0.1792 ± 0.1475	0.0086 ± 0.0098	0.0255 ± 0.0257	0.1167 ± 0.1169
Infomap	0.0283 ± 0.0167	0.0865 ± 0.0490	0.2995 ± 0.1903	0.014 ± 0.0145	0.0407 ± 0.0393	0.2248 ± 0.1423
Label Prop	0.0151 ± 0.0077	0.0394 ± 0.0181	0.2689 ± 0.1696	0.0072 ± 0.0066	0.0216 ± 0.0147	0.2089 ± 0.1412
Louvain	0.0160 ± 0.0154	0.0464 ± 0.0444	0.1235 ± 0.1294	0.0098 ± 0.0111	0.0288 ± 0.0298	0.1141 ± 0.1166
Walktrap	0.0280 ± 0.0166	0.0809 ± 0.0436	0.2891 ± 0.1895	0.0140 ± 0.0136	0.0414 ± 0.0347	0.1979 ± 0.1418

- In the FAO dataset, TF scores count the most frequent words and are able to identify the most critical words in each document.
- In the case of structured text (CiteULike), we observe that the GoW representation performs better than the simple statistical term frequency scores.
- Given the GoW representation, we observe that when N=3 the results are better than N=2, where N is the number of successive words that are linked to any word. However, the linking of more words than N=3 successive words, makes the graph of words almost complete, so centralities become identical and the graph has only one community (all the graph).
- Among the centrality measures, closeness centrality performs better than the other measures. In the case of N=2, Mapping Entropy Betweenness centrality has larger Jaccard index than all other methods.
- Among the community detection approaches, the Infomap communities contain the most important words on average and therefore obtain higher Jaccard, Average Precision and P@10.
- Community detection approaches are not superior to centrality scores, in all cases examined.
- Our proposed Mapping Entropy Closeness (MEC) centrality measure is the second most performing keyword extraction approach, in the case of Jaccard index, following the Mapping Entropy Betweenness (MEB) scores.

REFERENCES

Abilhoa, W.D., De Castro, L.N.: A keyword extraction method from twitter messages represented as graphs. Applied Mathematics and Computation 240 (2014) 308-325

Beliga, S., Mestrovic, A., Martincic-Ipsic, S.: An overview of graph-based keyword extraction methods and approaches. Journal of information and organizational sciences 39(1) (2015) 1-20

Blondel, V.D., Guillaume, J.L., Lambiotte, R., Lefebvre, E.: Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment 2008(10) (2008) P10008

Boudin, F.: A comparison of centrality measures for graph-based keyphrase extraction. In: International Joint Conference on Natural Language Processing (IJC-NLP). (2013) 834-838

Clauset, A., Newman, M.E.J., Moore, C.: Finding community structure in very large networks. Physical Review E 70(6) (2004) 066111

Gialampoukidis, I., Kalpakis, G., Tsikrika, T., Vrochidis, S., Kompatsiaris, I.: Key-player identification in terrorism-related social media networks using centrality measures. In: European Intelligence and Security Informatics Conference (EISIC 2016), August. (2016) 17-19

Grineva, M., Grinev, M., Lizorkin, D.: Extracting key terms from noisy and multitheme documents. In: Proceedings of the 18th international conference on World wide web, ACM (2009) 661-670

Lahiri, S., Choudhury, S.R., Caragea, C.: Keyword and keyphrase extraction using centrality measures on collocation networks. arXiv preprint arXiv:1401.6571 (2014)

Nie, T., Guo, Z., Zhao, K., Lu, Z.M.: Using mapping entropy to identify node centrality in complex networks. Physica A: Statistical Mechanics and its Applications 453 (2016) 290-297

Pons, P., Latapy, M.: Computing communities in large networks using random walks. Journal of Graph Algorithms and Applications 10(2) (2006) 191-218

Raghavan, U.N., Albert, R., Kumara, S.: Near linear time algorithm to detect community structures in large-scale networks. Physical Review E 76(3)

Rosvall, M., Bergstrom, C.T.: Maps of random walks on complex networks reveal community structure. Proceedings of the National Academy of Sciences 105(4) (2008) 1118-1123

Rousseau, F., Vazirgiannis, M.: Graph-of-word and tw-idf: new approach to ad hoc ir. In: Proceedings of the 22nd ACM international conference on Information & Knowledge Management, ACM (2013) 59-68

Tsatsaronis, G., Varlamis, I., Nrvag, K.: Semanticrank: ranking keywords and sentences using semantic graphs. In: Proceedings of the 23rd International Conference on Computational Linguistics, Association for Computational Linguistics (2010) 1074-1082

Xie, Z.: Centrality measures in text mining: prediction of noun phrases that appear in abstracts. In: Proceedings of the ACL student research workshop, Association for Computational Linguistics (2005) 103-108

THANK YOU!!!

KRISTINA
A Knowledge-Based Information Agent with Social Competence and Human Interaction Capabilities

TENSOR

This work was supported by the projects H2020-645012 (KRISTINA) and H2020-700024 (TENSOR), funded by the European Commission.