Motivation

Graph-degeneracy is better than PageRank for keyword extraction [Rousseau & Vazirgiannis 2015], but:

- retaining only the main core is suboptimal: one cannot expect all the keywords to live in the top level of the hierarchy → how to automatically select the best hierarchy level?
- dens: go down the hierarchy until a drop in density is observed
- inf: go down the hierarchy as long as the shells ↗ in size
- working with subgraphs lacks flexibility → how to rank nodes individually while retaining the valuable cohesiveness information captured by degeneracy?
- CoreRank (CR): (1) assign to each node the sum of the core or truss numbers of its neighbors. (2) select the elbow in the scores curve (CRE) or retain the top p% nodes (CRP)

Graph degeneracy

k-CORE DECOMPOSITION
- the k-core of G = (V, E) is a maximal connected subgraph of G in which every vertex \( v \) has at least degree \( k \) [Seidman 1983]
- \( v \) has core number \( k \) if it belongs to the \( k \)-core but not to the \((k + 1)\)-core
- the k-core decomposition of \( G \) is the set of all its cores from \( k = 0 \) (G itself) to \( k = K_{\text{max}} \) (its main core)
- complexity: \( O(n + m) \) resp. \( O(n \log(n)) \) in time in the (un)weighted cases, \( O(n) \) in space [Batagelj & Zaveršnik 2002]

K-TRUSS DECOMPOSITION
- the K-truss of \( G = (V, E) \) is its largest subgraph where every edge \( e \) belongs to at least \( K \) = 2 triangles [Cohen 2008]
- \( e \) has truss number \( K \) if it belongs to the \( K \)-truss but not to the \((K + 1)\)-truss
- the truss number of \( v \) is the maximum truss number of its adjacent edges
- the K-truss decomposition of \( G \) is the set of all its \( K \)-trusses from \( 2 \) (G) to \( K_{\text{max}} \)
- complexity: \( O(m^2) \) in time and \( O(m + n) \) in space [Wang & Cheng 2012]

Degeneracy and Spreading Influence

- in social networks, the best spreaders are not the highly connected individuals, but those located at the core of the network [Kitsak 2010]
- the truss number is an even better indicator of spreading influence than the core number [Malliaros et al. 2016]
- the spreading influence of a node is related to its structural position within the graph (density and cohesiveness) rather than to its prestige (random walk-based degree) ⇒ influential words should make better keywords

Datasets

- Huith2003: 500 abstracts from the Inspec physics & engineering database
- Marujo2012: 450 web news stories covering 10 different topics
- Semeval: 100 scientific papers from the ACM

Empirical Methods on Natural Language Processing (EMNLP) 2016, Austin, TX, USA