

# Normalized Entity Graph for Computing Local Coherence

Mohsen Mesgar and Michael Strube

Heidelberg Institute for Theoretical Studies gGmbH  
Schloss-Wolfsbrunnenweg 35  
69118 Heidelberg, Germany

(mohsen.mesgar|michael.strube)@h-its.org

## Abstract

Guinaudeau and Strube (2013) introduce a graph based model to compute local entity coherence. We propose a computationally efficient normalization method for these graphs and then evaluate it on three tasks: sentence ordering, summary coherence rating and readability assessment. In all tasks normalization improves the results.

## 1 Introduction

Guinaudeau and Strube (2013) introduce a graph based model (henceforth called *entity graph*) to compute local entity coherence. Despite being unsupervised, the entity graph performs on par with Barzilay and Lapata’s (2005; 2008) supervised entity grid on the tasks of sentence ordering, summary coherence rating and readability assessment. The entity graph also overcomes shortcomings of the entity grid with regard to computational complexity, data sparsity and domain dependence.

The entity graph is a bipartite graph where one set of nodes represents entities and the other set of nodes represents the sentences of a document. Guinaudeau and Strube (2013) apply a one mode projection on sentence nodes (Newman, 2010) and then compute the average out-degree of sentence nodes to determine how coherent a document is. They describe variants of their entity graph which take the number of shared entities between sentences and their grammatical functions into account thus resulting in weighted bipartite graphs and weighted one mode projections. Here, we propose to normalize weights for the entity graph. Normalization allows to include distance between mentions of the same entity, which improves the performance on all three tasks thus confirming research in related areas which states that normalizing weights leads to better performance (Zhou et al., 2008; Zweig and Kaufmann, 2011).

## 2 The Entity Graph

The entity graph (Guinaudeau and Strube, 2013),  $G = (V, E)$ , represents the relations between sentences and entities in a text, where node set  $V$  contains all sentences and entities in a text and  $E$  is the set of all edges between sentences and entities. Let function  $w(s_i, e_j)$  indicate the weight of an edge which connects sentence  $s_i$  and entity  $e_j$ . If  $w(s_i, e_j) = 1$ , then this edge indicates that there is a mention of  $e_j$  in sentence  $s_i$ . In order to realize the insight from Grosz et al. (1995) that certain syntactic roles are more important than others, the syntactic role of  $e_j$  in  $s_i$  can be mapped to an integer value (Guinaudeau and Strube, 2013):

$$w(s_i, e_j) = \begin{cases} 3 & \text{if } e_j \text{ is subject in } s_i \\ 2 & \text{if } e_j \text{ is object in } s_i \\ 1 & \text{otherwise} \end{cases}$$

Figure 1 illustrates a weighted entity graph for three sentences.

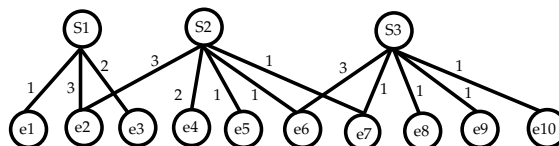


Figure 1: Weighted entity graph

Three types of one-mode projections capture relations between sentences,  $P_U$ ,  $P_W$  and  $P_{Acc}$ .  $P_U$  creates an edge between two sentences if they share at least one entity.  $P_W$  captures the intuition that the connection between two sentences is stronger the more entities they share by means of weighted edges, where the weights equal the number of entities shared by sentences (Newman, 2004). The third type of projection,  $P_{Acc}$ , integrates syntactic information in the edge weights calculated by the following formula:

$$W_{ik} = \sum_{e \in E_{ik}} w(e, s_i) \cdot w(e, s_k).$$

Figure 2 shows the three kinds of one-mode projections used in the entity graph.

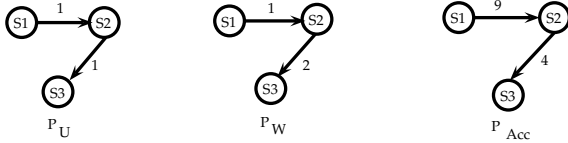


Figure 2: One-mode projections

While the entity grid (Barzilay and Lapata, 2008) uses information about sentences which do not share entities by means of the “-” transition, the entity graph cannot employ this negative information. Here, we propose a normalization for the entity graph and its corresponding one-mode projections which is based on the *relative* importance of entities and, in turn, the *relative* importance of sentences. Including negative information allows to normalize the importance of entities according to sentence length (measured in terms of entity mentions), and hence to capture distance information between mentions of the same entity. This brings the entity graph closer to Stoddard’s (1991, p.30) notion of cohesion: “The relative cohesiveness of a text depends on the number of cohesive ties [...] and on the distance between the nodes and their associated cohesive elements.” By using this information, edge weights are set less arbitrary which leads to the more sound method and higher performance in all tasks.

### 3 Normalized Entity Graph

The entity graph weighs edges by the number of entities sentences share ( $P_W$ ) and which syntactic functions the entities occupy ( $P_{Acc}$ ). Here we normalize the weights by the number of entities in a sentence. This takes negative information into account as entities which do not occur in other sentences also count. Hence normalization captures the relative importance of entities as well as the relative importance of sentences.

We follow Newman (2004) by applying node degree normalization. For  $P_W$ , we divide the weight of each edge by the degree of the corresponding sentence node. If a sentence contains many entities, then the amount of information each entity contributes is reduced. Assume  $\|s_i\|$  as the number of entities in sentence  $s_i$ . The importance of entity  $e_j$  for  $s_i$  is

$$Imp(s_i, e_j) = \frac{1}{\|s_i\|}.$$

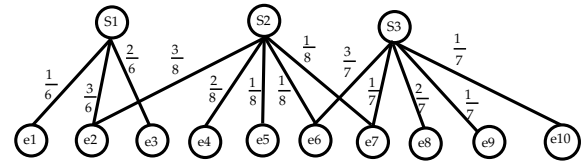


Figure 3: Normalized entity graph

For  $P_{Acc}$  we divide the weight of each edge by the sum of all edges’ weights of a sentence. This gives the importance of each entity in a sentence relative to the sentence’s other entities (see Figure 3).

$$Imp(s_i, e_j) = \frac{w(s_i, e_j)}{\sum_{e \in Entities} w(s_i, e_e)}.$$

For also normalizing the one-mode projection we introduce a virtual node  $TC$  capturing the textual content of all sentences (inspired by the graph based information retrieval model of Rode (2008)). The virtual node  $TC$  is connected to all sentences (see Figure 4).

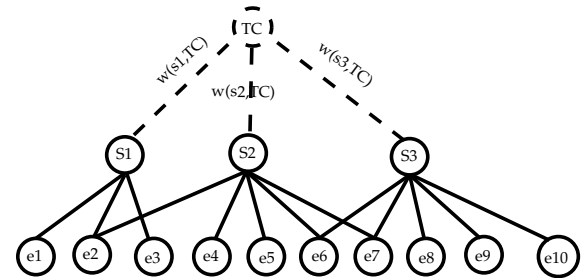


Figure 4: Entity graph with virtual node

Rode (2008) uses the following formula to compute weights on the edges between the sentence nodes and  $TC$ :

$$w(s_i, TC) = \frac{Score(s_i|TC)}{\sum_{s_t} Score(s_t|TC)},$$

where the function  $Score(s_i|TC)$  is the number of entities in  $s_i$  which have overlap with  $TC$ . This value is equal to the degree of each sentence.

Since we are interested in local coherence, we restrict  $TC$  to pairs of sentences (See Figure 5). Subsequently, instead of  $w(s_i, TC)$ , we use the notation  $lw_{s_i}^{s_j}$  (local weight of sentence  $s_i$  according to sentence  $s_j$ ).

We define the normalized one-mode projection as follows:

$$W_{s_{ij}} = \sum_{e \in E_{s_{ij}}} \left\{ (lw_{s_i}^{s_j} \cdot Imp(s_i, e)) + (lw_{s_j}^{s_i} \cdot Imp(s_j, e)) \right\}.$$

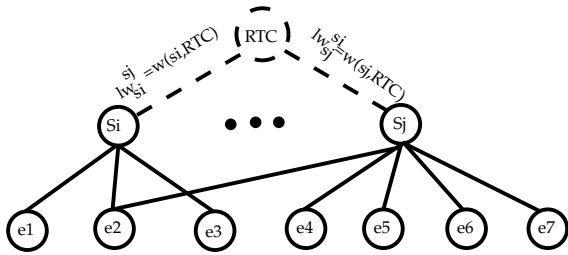


Figure 5: Restricted TC for a pair of sentences

Similar to Rode (2008), we use the product of  $lw_{s_i}^{s_j}$  and  $Imp(s_i, e)$  to approximate the salience of entity  $e$  in sentence  $s_i$ . This prevents the model to get biased by the length of sentences.

This method can be applied to graphs with edges weighted according to syntactic role ( $P_{Acc}$ ). To compute the connection’s strength of a pair of sentences we follow Yang and Knoke’s (2001) approach: The path length in a weighted graph is the sum of the edge weights in the path. In our case, each path is defined between a pair of sentences of the entity graph, so the number of edges of all paths are equal to two. Figure 6 shows the normalized projections where the weights have been computed by the above formula.

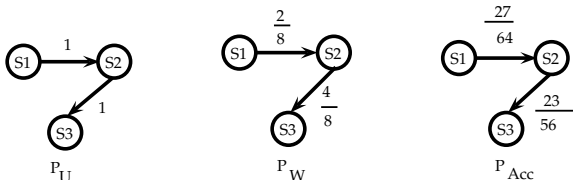


Figure 6: Normalized projections

## 4 Experiments

We compare the normalized entity graph with the entity graph on all tasks, Guinaudeau and Strube (2013) compared their work with the entity grid (Barzilay and Lapata, 2008; Elsner and Charniak, 2011): sentence ordering, summary coherence rating and readability assessment. Following Guinaudeau and Strube (2013) we test statistical significance with the Student’s t-test and Bonferroni correction, to check whether the best result (bold value in the tables) is significantly different from the results of the entity graph and the normalized entity graph. Diacritics \*\* indicate significance level 0.01, \* indicates significance level 0.05.

	Acc	F
Random	0.496	0.496
B&L	0.877	0.877
E&C	0.915	0.915
Entity graph, G&S		
$P_U, Dist$	0.830	0.830**
$P_W, Dist$	0.871	0.871
$P_{Acc}, Dist$	0.889	0.889
Normalized entity graph		
$P_U, Dist$	0.830	0.830**
$P_W, Dist$	0.886	0.886
$P_{Acc}, Dist$	<b>0.909</b>	<b>0.909</b>

Table 1: Discrimination, baselines and entity graph vs. normalized entity graph

### 4.1 Sentence Ordering

This task consists of two subtasks: discrimination and insertion. In both subtasks we evaluate whether our model can distinguish between the correct order of sentences in a document and an incorrect one. Experimental setup and data follow Guinaudeau and Strube (2013) (61 documents from the English test part of the CoNLL 2012 shared task (Pradhan et al., 2012)).

For discrimination we use 20 permutations of each text. Table 1 shows the results. Results for Guinaudeau and Strube (2013), G&S, are reproduced, results for Barzilay and Lapata (2008), B&L, and Elsner and Charniak (2011), E&C, were reproduced by Guinaudeau and Strube (2013).

The unweighted graph,  $P_U$ , does not need normalization. Hence the results for the entity graph and the normalized entity graph are identical. Normalization improves the results for the weighted graphs  $P_W$  and  $P_{Acc}$  with  $P_{Acc}$  outperforming B&L considerably and closely approaching E&L.

Sentence insertion is more difficult than discrimination. Following Elsner and Charniak (2011), we use two measures for evaluation: Accuracy (Acc.) and the average proportion of correct insertions per document (Ins.).

	Acc.	Ins.
Random	0.028	0.071
E&C	0.068	0.167
Entity graph, G&S		
$P_U, Dist$	0.062**	0.101**
$P_W, Dist$	0.075	0.114**
$P_{Acc}, Dist$	0.071	0.102**
Normalized entity graph		
$P_U, Dist$	0.062**	0.101**
$P_W, Dist$	<b>0.085</b>	<b>0.154</b>
$P_{Acc}, Dist$	0.077	0.157

Table 2: Insertion, baselines and entity graph vs. normalized entity graph

	Acc.	F
B&L	0.833	
	Entity graph, G&S	
$P_U$	0.800	0.815
$P_W$	0.613	0.613*
$P_{Acc}$	0.700	0.704
	Normalized entity graph	
$P_U$	<b>0.800</b>	<b>0.815</b>
$P_W$	0.775	0.775
$P_{Acc}$	0.788	0.788

Table 3: Summary Coherence Rating, B&L and entity graph vs. normalized entity graph

Table 2 shows that the normalized entity graph outperforms the entity graph for  $P_W$  and  $P_{Acc}$  (again, no difference for  $P_U$ ). The normalized entity graph outperforms E&C in Acc. and approaches it in Ins. The high value for Ins. shows that if the normalized entity graph makes false decisions they are closer to the original ordering than the mistakes of the entity graph.

## 4.2 Summary Coherence Rating

We follow Barzilay and Lapata (2008) for evaluating whether the normalized entity graph can decide whether automatic or human summaries are more coherent (80 pairs of summaries extracted from DUC 2003). Human coherence scores are associated with each pair of summarized documents (Barzilay and Lapata, 2008).

Table 3 displays reported results of *B&L* and reproduced results of the entity graph and our normalized entity graph. Normalizing significantly improves the results for  $P_W$  and  $P_{Acc}$ .  $P_U$  is still slightly better than both, but in contrast to the entity graph, this difference is not statistically significant. We believe that better weighting schemes based on linguistic insights eventually will outperform  $P_U$  and B&L (left for future work). Distance information always degrades the results for this task (see Guinaudeau and Strube (2013)).

## 4.3 Readability Assessment

Readability assessment aims to distinguish texts which are difficult to read from texts which are easier to read. In experiments, Barzilay and Lapata (2008) assume that articles taken from Encyclopedia Britannica are more difficult to read (less coherent) than the corresponding articles from Encyclopedia Britannica Elementary, its version for children. We follow them with regard to data (107 article pairs), experimental setup and evaluation.

Table 4 compares reported results by Schwarm

	Acc.	F
S&O	0.786	
B&L	0.509	
B&L + S&O	0.888	
	Entity graph, G&S	
$P_U, Dist$	0.589	0.589**
$P_W, Dist$	0.570	0.570**
$P_{Acc}, Dist$	0.766	0.766**
	Normalized entity graph	
$P_U, Dist$	0.589	0.589**
$P_W, Dist$	<b>0.897</b>	<b>0.897</b>
$P_{Acc}, Dist$	0.850	0.850

Table 4: Readability assessment, baselines and entity graph vs. normalized entity graph

and Ostendorf (2005), S&O, Barzilay and Lapata (2008), B&L, a combined method, B&L + S&O, reproduced results for the entity graph, G&S, and our normalized entity graph. Distance information always improves the results.

Sentences in the *Britannica Elementary* are simpler and shorter than in the *Encyclopedia Britannica*. The entity graph does not take into account the effect of entities not shared between sentences while the normalized entity graph assigns a lower weight if there are more of these entities. Hence, *Britannica Elementary* receives a higher cohesion score than *Encyclopedia Britannica* in our model. Adding grammatical information, does not help, because of the influence of the number of entities (shared and not shared) outweighs the influence of syntactic roles. The normalized entity graph ( $P_W, Dist$ ) does not only outperform the entity graph (significantly) and B&L but also S&O and the combination B&L + S&O.

## 5 Conclusion

We proposed a normalization method for the entity graph (Guinaudeau and Strube, 2013). We compared our model to the entity graph and to the entity grid (Barzilay and Lapata, 2008) and showed that normalization improves the results significantly in most tasks. Future work will include adding more linguistic information, stronger weighting schemes and application to other readability datasets (Pitler and Nenkova, 2008; De Clercq et al., 2014).

## Acknowledgments

This work has been funded by the Klaus Tschira Foundation, Heidelberg, Germany. The first author has been supported by a Heidelberg Institute for Theoretical Studies Ph.D. scholarship.

## References

- Regina Barzilay and Mirella Lapata. 2005. Modeling local coherence: An entity-based approach. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics*, Ann Arbor, Mich., 25–30 June 2005, pages 141–148.
- Regina Barzilay and Mirella Lapata. 2008. Modeling local coherence: An entity-based approach. *Computational Linguistics*, 34(1):1–34.
- Orphée De Clercq, Véronique Hoste, Bart Desmet, Philip Van Oosten, Martine De Cock, and Lieve Macken. 2014. Using the crowd for readability prediction. *Natural Language Engineering*, 20(3):293–325.
- Micha Elsner and Eugene Charniak. 2011. Extending the entity grid with entity-specific features. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, Portland, Oreg., 19–24 June 2011, pages 125–129.
- Barbara J. Grosz, Aravind K. Joshi, and Scott Weinstein. 1995. Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, 21(2):203–225.
- Camille Guinaudeau and Michael Strube. 2013. Graph-based local coherence modeling. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Sofia, Bulgaria, 4–9 August 2013, pages 93–103.
- Mark E.J. Newman. 2004. Analysis of weighted networks. *Physical Review E*, 70(5):056131.
- Mark E.J. Newman. 2010. *Networks: An Introduction*. Oxford University Press, New York, N.Y.
- Emily Pitler and Ani Nenkova. 2008. Revisiting readability: A unified framework for predicting text quality. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, Waikiki, Honolulu, Hawaii, 25–27 October 2008, pages 186–195.
- Sameer Pradhan, Alessandro Moschitti, Nianwen Xue, Olga Uryupina, and Yuchen Zhang. 2012. CoNLL-2012 Shared Task: Modeling multilingual unrestricted coreference in OntoNotes. In *Proceedings of the Shared Task of the 16th Conference on Computational Natural Language Learning*, Jeju Island, Korea, 12–14 July 2012, pages 1–40.
- Henning Rode. 2008. *From document to entity retrieval: Improving precision and performance of focused text search*. Ph.D. thesis, Enschede, June.
- Sarah E. Schwarm and Mari Ostendorf. 2005. Reading level assessment using support vector machines and statistical language models. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics*, Ann Arbor, Mich., 25–30 June 2005, pages 523–530.
- Sally Stoddard. 1991. *Text and Texture: Patterns of Cohesion*. Ablex, Norwood, N.J.
- Song Yang and David Knoke. 2001. Optimal connections: Strength and distance in valued graphs. *Social networks*, 23(4):285–295.
- Tao Zhou, Jie Ren, Matúš Medo, and Yi-Cheng Zhang. 2008. Bipartite network projection and personal recommendation. *Physical Review E*, 76(4). 046115.
- Katharina A. Zweig and Michael Kaufmann. 2011. A systematic approach to the one-mode projection of bipartite graphs. *Social Network Analysis and Mining*, 1:187–218.