

Measuring Narrative Fluency by Analyzing Dynamic Interaction Networks in Textual Narratives

O-Joun Lee

Future IT Innovation Laboratory
Pohang Univ. of Science and Technology
Pohang-si, Republic of Korea 37673
ojlee112358@postech.ac.kr

Jin-Taek Kim*

Future IT Innovation Laboratory
Pohang Univ. of Science and Technology
Pohang-si, Republic of Korea 37673
jintaek@postech.ac.kr

Abstract

This study aims to assess the fluency of narratives in textual multimedia (e.g., news articles, academic publications, novels, etc.). We measure the narrative fluency based on whether relationships between entities in the narrative (i.e., subjects and objects of events that compose the narrative) are consistently described with adequate rapidity. The relationships are represented by a dynamic interaction network (called ‘entity network’), which has entities as nodes and co-occurrences between the entities as edges. Lack of consistency makes users confused about what the textual narratives want to present. If a narrative consistently concentrates on a topic or subject, its entity network will have few entities with high node centrality. Using consistency of the high centrality entities, we assess the fluency with three criteria: (i) consistency in each paragraph, (ii) consistency in the overall narrative, and (iii) consistency between the title and body. The rapidity of narrative development has to be appropriate for expected readers of the textual narratives. Too low rapidity causes redundancy, and high rapidity hinders the understandability of the narratives. We assume structural changes in the entity network reflect the narrative rapidity. The structural change is measured by embedding structures of the entity network. Finally, we evaluated the effectiveness of the proposed methods using the editorials of the New York Times and human evaluators.

1 Introduction

Recently, various studies have attempted to quantitatively measure what has been qualitatively assessed based on human intuition and experience, such as story similarity [LJJ18, LJ19b, LJ20], creativity [ES15], trustworthiness [LNJ⁺17], and so on. These studies have mainly been conducted on interdisciplinary areas between computer science and the humanities/social science. They attempt to make subjects of the humanities/social science computational.

As one of these attempts, this study aims to quantify the fluency of narratives, which is one of the significant factors for evaluating writings [Huc15]. Narratives are the most fundamental media for exchanging information between human beings. According to Lakoff and Narayanan [LN10], “Narratives structure our understanding of

* Corresponding author.

the world and of ourselves.” Therefore, assessing the quality of narratives is significant not only for multimedia content analysis and its applications but also for human-computer interaction. Taghipour and Ng [TN16] also attempted to score essays automatically by using the convolutional recurrent neural network. However, their method cannot be used as a standard indicator of the narrative quality, which should always assign the same score on a narrative.

Various studies [BGL⁺19, LB19, LJ19a, WCW09] have applied interaction networks between characters (character networks) for analyzing fictional narratives (stories). We extend this model, which has only been applied to fictional and artistic narratives, to cover general narratives, including news articles or academic publications. The character network foots on the assumption that interactions between characters compose fictional narratives. A narrative is a series of events, and an event consists of interactions between characters [McK97, McK16]. However, general narratives do not only depict relationships between personified entities (characters).

For example, a history book can describe interactions between nations or other social organizations, and research articles depict relationships among abstractive concepts. To interpret what the narratives accurately attempt to describe, we have to analyze the meanings of each interaction and relationship. However, even though we do not know the meanings of relationships, we can analyze how the entire relationships between entities are gradually presented or explained. This point is as with that our previous studies [LJJ18, LJ19b] have attempted to analyze the narrative development by only using frequencies of interactions between characters. This approach also enables us to apply the proposed methods on various kinds of media without significant modification, while the existing studies [SLE15, SMS15] measured the fluency based on the domain knowledge.

First, we have to define the entities and their interactions. Similar to the existing narrative models [CR17, MAW⁺18], we define entities as subjects and objects of each interaction. In video or audio, finding interactions and entities involved in the interactions is abstruse. Thus, as a preliminary study, we restrict our research subjects into textual narratives, e.g., news articles, academic publications, non-fiction books, novels, essays, etc. Each sentence in the text is used as a unit of interactions. And, entities correspond to nouns (or noun phrases), which can be subjects or objects of the sentence. The entity and the interaction can be defined as follows;

Definition 1 (Entity and Interaction) *Suppose that S is a set of sentences in a textual narrative, \mathbb{T} . When $s_i \in S$ is the i -th sentence in S , s_i also corresponds to the i -th interaction between entities. E , which is a set of entities in \mathbb{T} , consists of nouns and noun phrases that appeared in sentences within S . If two entities (e_a and e_b) co-occur in s_i , we can assume that s_i describes a relationship between e_a and e_b , even though we do not know meanings of s_i .*

The narrative is time-sequential. Thus, the existing studies [Bos16, LJ19a] segmented narratives into logical and regular units, such as scenes. A scene is defined as a period that does not contain changes in spatio-temporal backgrounds [McK97]. Each scene describes a concluded event within a background. However, the general narratives are far more diverse than the fictional ones. Interactions in the general narrative can be segmented into events, but they do not always have distinguishable backgrounds. Thus, we employ paragraphs as a unit of events, since paragraphs in well-written texts usually have topical coherence and completion. By using the paragraph as a time window, we define a dynamic interaction network between entities appeared in a narrative as follows;

Definition 2 (Entity Network) *Suppose that $|E|$ is the number of entities that appeared in a narrative, \mathbb{T} . When $N(\mathbb{T})$ indicates an entity network of \mathbb{T} , $N(\mathbb{T})$ can be defined as a matrix $\in \mathbb{R}^{|E| \times |E|}$. Each component of $N(\mathbb{T})$ means relationships between two entities. By defining $N(\cdot)$ in each paragraph, we can observe the development of the relationships. When P is a set of paragraphs in \mathbb{T} , and p_l indicates the l -th paragraph, $N(p_l)$ indicates an entity network for p_l . This can be formulated as:*

$$N(\mathbb{T}) = \sum_{l=1}^{|P|} N(p_l) = \begin{bmatrix} f_{1,1} & \cdots & f_{1,N} \\ \vdots & \ddots & \vdots \\ f_{N,1} & \cdots & f_{N,N} \end{bmatrix}, \quad (1)$$

where $f_{i,j}$ indicates frequency of interactions between e_i and e_j . We measure $f_{i,j}$ using the number of sentences that e_i and e_j co-occurred.

We measure the narrative fluency based on the entity network and the following two assumptions. First, the topical coherence of a paragraph will be exposed by the centrality of its keywords on the entity network. Thus,

within a paragraph, there should be few entities with significantly higher centrality than the other entities. If a narrative consistently focuses on a topic, the keywords will also be consistent in the overall narrative (RQ 1). Second, the relationships between entities have to be described in an appropriate rapidity to deliver the relationships to users understandably. If we use too few interactions or events for depicting content, there can be logical leaps. On the other hand, if we describe the content too slowly, there might be meaningless redundancy. Therefore, the narrative should have an adequate rapidity of its development regarding its purposes and expected readers. The narrative development will accompany new entities and new relationships between the entities. Thus, we assume that the rapidity of narrative development can be measured by structural changes in entity networks (RQ 2).

2 Measuring Narrative Fluency

This section first describes the way how we have composed entity networks, briefly. Then, we present the proposed method for measuring narrative fluency with the two criteria: (i) the narrative consistency and (ii) the rapidity of the narrative development.

2.1 Composing Entity Networks

We collected 20 recent editorials published in the New York Times¹. Titles, headlines, and bodies of the editorials were collected and preprocessed by using the NLTK library of Python². We conducted tokenization, stemming, and parts of speech (POS) tagging for the collected texts. Then, we annotated occurrences and co-occurrences of only nouns and pronouns, which are tagged as ‘NN,’ ‘NNS,’ ‘NNP,’ or ‘NNPS’ by the POS tagger in the NLTK library. These nouns and pronouns are entities. We composed entity networks based on occurrences and co-occurrences of the entities in each sentence and paragraph. To segment sentences, we used the capitalization, punctuation marks, and dictionary for frequently-used acronyms (e.g., Mr., Ms., etc.). Also, the entity network includes cyclic edges (e.g., $f_{a,a}$ in Eq. 1) to represent the occurrence frequency of entities.

2.2 Measuring Narrative Consistency of Paragraphs

We measure the narrative consistency with two viewpoints. First, each paragraph has to focus on one topic. Thus, there should be few entities that have far higher centrality than the other entities. The entities will be keywords that represent the topic. Therefore, we measure three well-known node centrality measurements (e.g., degree, betweenness, closeness centrality) for each entity on entity networks. The centrality measurements are normalized into $[0, 1]$ and aggregated by the arithmetic mean. We use the arithmetic mean as the centrality of each keyword. However, this way cannot consider which kinds of centrality are more significant for the narrative fluency. In future studies, we will compare the significance of the centrality measurements by applying weighting factors to them. We assess the narrative consistency in each paragraph by using the entropy of the centrality of entities. This can be formulated as:

$$C^{all}(p_l) = \frac{1}{|E_l|} \times \sum_{\forall e_a \in E_l} -\log \mathcal{C}_l(e_a), \quad (2)$$

where $E_l \subset E$ indicates a set of entities that appeared in p_l , and $\mathcal{C}_l(e_a)$ refers to centrality of e_a on $N(p_l)$. $C^{all}(p_l)$ measures consistency of a paragraph. For the entire textual narrative, we aggregate the consistency of paragraphs as: $C^{all}(\mathbb{T}) = \frac{1}{|P|} \times \sum_{\forall p_l \in P} C^{all}(p_l)$.

Second, keywords of a narrative should have high centrality on overall paragraphs in the narrative. Similar to the previous one, we assess whether keywords have consistently high centrality, based on the entropy. This can be formulated as:

$$C^{key}(\mathbb{T}) = \left[1 + \frac{1}{|P| \cdot |K|} \times \sum_{\forall p_l \in P} \sum_{\forall e_a \in K} -\log \mathcal{C}_l(e_a) \right]^{-1}, \quad (3)$$

where $K \subset E$ is a set of keywords of \mathbb{T} . We compose K by clustering entities into two clusters according to their centrality by using k -means clustering with two initial centroids: maximum and minimum centrality. Among

¹<https://www.nytimes.com/section/opinion/editorials>

²<http://www.nltk.org/>

the two clusters, we assume that elements in the cluster with higher centrality are keywords. Although using a threshold will be much simpler than the clustering, it cannot deal with the diversity of textual narratives.

Third, users expect topics of textual narratives from their titles and keywords annotated by creators. Since news articles are our experimental subjects, entities in their titles and headlines have to match with keywords discovered by using the entity network. Thus, we measure their concurrence based on the Jaccard index. This can be formulated as:

$$C^{title}(\mathbb{T}) = \frac{|(E_t \cup E_h) \cap K|}{|E_t \cup E_h \cup K|}, \quad (4)$$

where E_t and E_h are sets of entities that appeared in the title and headline of \mathbb{T} , respectively. We aggregate the three proposed measurements (C^{all} , C^{key} and C^{title}) by the arithmetic mean, after normalizing them into $[0, 1]$.

2.3 Measuring Rapidity of Narrative Development

We measure the rapidity of the narrative development by using structural changes in entity networks. To compare structures of entity networks, we represent the entity networks with a vector by using the graph embedding technique. We employ the Graph2Vec model [NCV+17], which aims to embed structures of graphs rather than characteristics of nodes or edges. Graph2Vec first extracts subgraphs from entity networks by using the WL (Weisfeiler-Lehman) relabeling process [SSvL+11]. Then, PV-DBOW in Doc2Vec [LM14] is applied to learn representations of the entity networks based on the composition of subgraphs. We notate a vector representation of $N(\mathbb{T})$ as $\Phi(N(\mathbb{T}))$.

We measure the rapidity by estimating how significant changes in the entity network are caused by a paragraph. Thus, we have to compare entity networks before and after the paragraph. The rapidity of narrative development on p_l is measured by the Euclidean distance between $\Phi(\sum_{i=1}^l N(p_i))$ and $\Phi(\sum_{i=1}^{l-1} N(p_i))$. This can be formulated as:

$$R(p_l) = \left\| \Phi \left(\sum_{i=1}^l N(p_i) \right) - \Phi \left(\sum_{i=1}^{l-1} N(p_i) \right) \right\|_2. \quad (5)$$

This study focuses on validating that we can measure the narrative fluency by analyzing the interaction network. We will sophisticatedly tune the proposed measurements in future studies (e.g., comparing the effectiveness of various distance metrics in measuring the narrative rapidity).

We assume that too slow or too fast changes in entity networks hinder the readability of the narrative. Therefore, we assess whether the rapidity is appropriate and consistent. After normalizing the rapidity of paragraphs into $[0, 1]$, we aggregate the difference between the optimal rapidity and rapidity on each paragraph. This can be formulated as:

$$R(\mathbb{T}) = \frac{1}{|P|} \times \sum_{\forall p_l \in P} |R(p_l) - \Theta_R|, \quad (6)$$

where Θ_R indicates the optimal rapidity. We attempt to search Θ_R through empirical experiments in the following section.

Additionally, the proposed measurements for the narrative consistency can be manipulated by splitting paragraphs more finely than the normal. We expect that too short paragraphs make $R(p_l)$ small. Thus, narrative consistency and rapidity will have a trade-off relationship.

3 Evaluation

To validate the research questions, we evaluated the accuracy of the measurements by estimating fluency of the editorials in the New York Times. We compared the results of the proposed measurements with responses of 37 human evaluators. The evaluator group consists of students and faculty members of Chung-Ang University and Pohang University of Science and Technology. We asked the evaluators to read three editorials that they chose from our corpus and answer the following questionnaire.

- Q1. What are the keywords of this editorial?

Table 1: Accuracy of extracting keywords by using the proposed method (centrality) and TF-IDF.

Metric	Precision	Recall	F_1 measure
Centrality	0.86	0.66	0.75
TF-IDF	0.83	0.57	0.68

Table 2: Accuracy of detecting non-fluent paragraphs by using the narrative consistency and rapidity.

Metric	Precision	Recall	F_1 measure
Consistency	0.71	0.87	0.78
Rapidity	0.62	0.59	0.61

- Q2. Is this editorial consistently describing its topic? Please, answer in five degrees: very inconsistent, inconsistent, normal, consistent, and very consistent.
- Q3. If this editorial is inconsistent, please check paragraphs causing the inconsistency.
- Q4. How rapid is the narrative development of this editorial? Please, answer in five degrees: prolonged, slow, appropriate, fast, and very fast.
- Q5. If the rapidity of narrative development in this editorial is inadequate, please check paragraphs that causes the inappropriateness. Also, please annotate whether the paragraphs are redundant or unexpected.
- Q6. Is this editorial fluent? Please, answer in five degrees: very non-fluent, non-fluent, normal, fluent, and very fluent.
- Q7. If this editorial is not fluent, please check paragraphs that causes the non-fluency.

For normalization, the five choices of Q2 and Q6 were replaced with 0.2, 0.4, 0.6, 0.8, and 1.0, respectively. Also, the choices of Q4 were transformed into 0.0, 0.5, 1.0, 0.5, and 0.0, respectively.

Based on the questionnaire, we conducted three experiments. First, if the entity network model is reasonable, high centrality entities will be keywords of each textual narrative. Thus, we compared keywords annotated by the evaluators (Q1) with automatically discovered ones. As a baseline method, we measured TF-IDF (Term Frequency-Inverse Document Frequency) of entities and clustered them according to the TF-IDF scores, similar to the proposed method. Accuracy of the keywords was assessed by the precision, recall, and F_1 measure.

The second and third rows of Table 1 show that both of the methods have high precision and low recall. To find its reason, we examined keywords that were not discovered by the centrality and TF-IDF. Most of the omitted keywords were called by various expressions, including pronouns and synonyms. For example, the following phrases can be used in similar meanings: U.S. government, American government, Federal government, Trump administration, Presidency of Donald Trump, Washington D.C., etc. This variety of expressions makes co-occurrence frequency of entities dispersed. Vocabulary diversity makes texts smooth and fluent, while it is a challenging issue for composing accurate entity networks. Also, the centrality exhibited higher accuracy for discovering keywords than TF-IDF. However, the amount of improvement was insignificant. Even if the entity network is independent of kinds of media, its performance has to be improved, considering the simplicity of TF-IDF.

Second, we validated RQ 1 and assessed the effectiveness of the narrative consistency measurements, based on Q2, Q3, Q6, and Q7. We examined correlations between (i) fluency annotated by the evaluators (Q6; FH), (ii) annotated consistency (Q2; CH), and (iii) automatically measured consistency (CA), using PCC (Pearson Correlation Coefficient). FH-CH and FH-CA verified RQ 1, and CH-CA and FH-CA exhibited the effectiveness of the measurements. Table 3 (a) presents the correlation coefficients.

Table 3: Correlations between the proposed measurements and the annotations of evaluators.

(a) Narrative Consistency				(b) Narrative Rapidity			
	FH	CH	CA		FH	RH	RA
FH	1.00	0.91	0.71	FH	1.00	0.66	0.74
CH	0.91	1.00	0.73	RH	0.66	1.00	0.62
CA	0.71	0.73	1.00	RA	0.74	0.62	1.00

On the experimental results, FH-CH was 0.91. Most of the evaluators gave the same scores for the fluency and consistency. FH-CA (0.71) was lower than FH-CH but still significant. Thus, the consistency was an essential factor of the narrative fluency. CH-CA (0.73) was lower than FH-CH but higher than FH-CA. This point indicates that the proposed measurement adequately reflected the consistency of the editorials.

Then, we compared inconsistent paragraphs annotated by the evaluators with ones detected by the proposed method. By modifying Eq. 3, we measured inconsistency of each paragraph as: $\sum_{e_a \in K} -\log \mathcal{C}_l(e_a)$. According to this metric, we sorted paragraphs in each editorial with a descending order. Paragraphs in the first quartile of the order were determined as the inconsistent ones. Accuracy for detecting the inconsistent paragraphs was assessed by the precision, recall, and F_1 measure.

As shown in the second row of Table 2, the proposed method exhibited high recall but low precision. Since we use keywords to measure the inconsistency, recognizing synonyms as individual entities might increase the inconsistency for paragraphs. Although the consistency showed reasonable performance overall, we have to find a better way of composing the entity network.

Finally, we validated RQ 2 and verified the effectiveness of the proposed measures for the rapidity of narrative development, based on Q4 to Q7. As with the previous one, we examined correlations between (i) fluency annotated by the evaluators (Q6; FH), (ii) annotated rapidity (Q4; RH), and (iii) automatically measured rapidity (RA). FH-RH and FH-RA verified RQ 1, and RH-RA and FH-RA exhibited the effectiveness of the rapidity measurement. Table 3 (b) presents the correlation coefficients.

FH-RH (0.66) was relatively lower than FH-CH. Also, FH-RA (0.74) was lower than FH-CA. These results mean the rapidity was less significant than the consistency to estimate the narrative fluency. One interesting point was that FH-RA was higher than RH-RA (0.62). The rapidity measurement was correlated to the narrative fluency but not much proportional to the rapidity of narrative development that the evaluators felt. The following experiment also showed this problem. Additionally, the proposed measurement exhibited the highest PCC for RH-RA on $\Theta_R = 0.45$. We searched the optimal Θ_R in $[0, 1]$ with a step size +0.05.

Also, we compared too fast and slow paragraphs annotated by the evaluators with ones detected using the consistency measurements. Using Eq. 5, we sorted the paragraphs in each editorial with descending order. Then, paragraphs in the first and fourth quartiles of the order are decided as too fast and slow paragraphs, respectively. Their accuracy was assessed by using the precision, recall, and F_1 measure.

Different from the consistency, precision and recall of the rapidity measurement were similar. However, as displayed in the third row of Table 2, accuracy for detecting abnormality on the rapidity was significantly lower than on the consistency. To find its reason, we have examined false positives and false negatives of the proposed method. Interestingly, the false positives were mostly on the beginning and ending parts of the editorials (maybe, introductions and conclusions), and most of the false negatives were on the middle parts of the editorials. These results indicate that the optimal rapidity of narrative development can be different according to the locations of paragraphs (or narrative time). The low PCC score for RH-RA could be affected by this problem, either.

4 Conclusion

We have proposed two kinds of measurements for assessing fluency of textual narratives. Also, their effectiveness was evaluated based on the editorials of the New York Times. However, this study has a few limitations. First, we could not conduct experiments on various kinds of textual narratives. Also, we assumed the optimal rapidity as a static value. Our further research will be focused on resolving these two problems.

Acknowledgements

This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the ICT Consilience Creative program (IITP-2019-2011-1-00783) supervised by the IITP (Institute for Information & communications Technology Planning & Evaluation).

References

- [BGL⁺19] Xavier Bost, Serigne Gueye, Vincent Labatut, Martha Larson, Georges Linarès, Damien Malinas, and Raphaël Roth. Remembering winter was coming. *Multimedia Tools and Applications*, September 2019. To Appear.
- [Bos16] Xavier Bost. *A storytelling machine?: automatic video summarization: the case of TV series*. PhD thesis, University of Avignon, France, November 2016.

- [CR17] Emmanouil Theofanis Chourdakis and Joshua Reiss. Constructing narrative using a generative model and continuous action policies. In *Proceedings of the Workshop on Computational Creativity in Natural Language Generation (CC-NLG@INLG 2017)*, pages 38–43, Santiago de Compostela, Spain, September 2017. Association for Computational Linguistics (ACL).
- [ES15] Ahmed M. Elgammal and Babak Saleh. Quantifying creativity in art networks. In Hannu Toivonen, Simon Colton, Michael Cook, and Dan Ventura, editors, *Proceedings of the 6th International Conference on Computational Creativity (ICCC 2015)*, pages 39–46, Park City, Utah, USA, June 2015. computationalcreativity.net.
- [Huc15] Geoffrey J. Huck. *What Is Good Writing?*, chapter Narrative Fluency, pages 102–124. Oxford University Press, September 2015.
- [LB19] Vincent Labatut and Xavier Bost. Extraction and analysis of fictional character networks: A survey. *ACM Computing Surveys*, 2019. To Appear.
- [LJ19a] O-Joun Lee and Jason J. Jung. Integrating character networks for extracting narratives from multimodal data. *Information Processing and Management*, 56(5):1894–1923, September 2019.
- [LJ19b] O-Joun Lee and Jason J. Jung. Modeling affective character network for story analytics. *Future Generation Computer Systems*, 92:458–478, March 2019.
- [LJ20] O-Joun Lee and Jason J. Jung. Story embedding: Learning distributed representations of stories based on character networks. *Artificial Intelligence*, 281:103235, April 2020.
- [LJJ18] O-Joun Lee, Nayoung Jo, and Jason J. Jung. Measuring character-based story similarity by analyzing movie scripts. In Alípio Mário Jorge, Ricardo Campos, Adam Jatowt, and Sérgio Nunes, editors, *Proceedings of the 1st Workshop on Narrative Extraction From Text (Text2Story 2018) co-located with the 40th European Conference on Information Retrieval (ECIR 2018)*, volume 2077 of *CEUR Workshop Proceedings*, pages 41–45, Grenoble, France, March 2018. CEUR-WS.org.
- [LM14] Quoc V. Le and Tomas Mikolov. Distributed representations of sentences and documents. In Eric P. Xing and Tony Jebara, editors, *Proceedings of the 31th International Conference on Machine Learning (ICML 2014)*, volume 32 of *JMLR Workshop and Conference Proceedings*, pages 1188–1196, Beijing, China, June 2014. JMLR.org.
- [LN10] George Lakoff and Srinu Narayanan. Toward a computational model of narrative. In *Proceedings of the 2010 AAAI Fall Symposium: Computational Models of Narrative*, volume FS-10-04 of *AAAI Technical Report*, pages 21–28, Arlington, VA, US, November 2010. AAAI.
- [LNJ+17] O-Joun Lee, Hoang Long Nguyen, Jai E. Jung, Tai-Won Um, and Hyun-Woo Lee. Towards ontological approach on trust-aware ambient services. *IEEE Access*, 5:1589–1599, February 2017.
- [MAW+18] Lara J. Martin, Prithviraj Ammanabrolu, Xinyu Wang, William Hancock, Shruti Singh, Brent Harrison, and Mark O. Riedl. Event representations for automated story generation with deep neural nets. In Sheila A. McIlraith and Kilian Q. Weinberger, editors, *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI 2018), the 30th innovative Applications of Artificial Intelligence (IAAI 2018), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI 2018)*, pages 868–875, New Orleans, Louisiana, USA, February 2018. AAAI Press.
- [McK97] Robert McKee. *Story: Substance, Structure, Style and the Principles of Screenwriting*. HarperCollins, New York, NY, USA, November 1997.
- [McK16] Robert McKee. *Dialogue: The Art of Verbal Action for Page, Stage, and Screen*. Twelve, July 2016.
- [NCV+17] Annamalai Narayanan, Mahinthan Chandramohan, Rajasekar Venkatesan, Lihui Chen, Yang Liu, and Shantanu Jaiswal. graph2vec: Learning distributed representations of graphs. *Computing Research Repository (CoRR)*, abs/1707.05005, July 2017.

- [SLE15] Oscar Saz, Yibin Lin, and Maxine Eskenazi. Measuring the impact of translation on the accuracy and fluency of vocabulary acquisition of english. *Computer Speech & Language*, 31(1):49–64, May 2015.
- [SMS15] Maryam Soleimani, Sima Modirkhamene, and Karim Sadeghi. Peer-mediated vs. individual writing: measuring fluency, complexity, and accuracy in writing. *Innovation in Language Learning and Teaching*, 11(1):86–100, June 2015.
- [SSvL⁺11] Nino Shervashidze, Pascal Schweitzer, Erik Jan van Leeuwen, Kurt Mehlhorn, and Karsten M. Borgwardt. Weisfeiler-lehman graph kernels. *Journal of Machine Learning Research*, 12:2539–2561, September 2011.
- [TN16] Kaveh Taghipour and Hwee Tou Ng. A neural approach to automated essay scoring. In Jian Su, Xavier Carreras, and Kevin Duh, editors, *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP 2016)*, pages 1882–1891, Austin, Texas, USA, November 2016. Association for Computational Linguistics.
- [WCW09] Chung-Yi Weng, Wei-Ta Chu, and Ja-Ling Wu. RoleNet: Movie analysis from the perspective of social networks. *IEEE Transactions on Multimedia*, 11(2):256–271, February 2009.