

Semantic Representations of Political Discourse: based on Topic Modeling and Algorithms

DAGIIMAA BALAAZ

School of language and culture,
University of the Humanities Ulaanbaatar, Mongolia

Abstract:- This research is a contribution to text network analysis using text mining algorithms, how the speaker creates discursive approaches of macrostructure and interfaces with knowledge based on the mental model within discourse macrostructure theory, developed by the Fairclough, Teun A. van Dijk, aimed to identify and analyze authors opinion and attitude in political speech. The detection of sentiment and emotional tone is also related to the method of semantic structure in discourse, which is a psychological factor. Therefore, the episodic and mental space, which study of the psychological factors based on the theory of knowledge, has practical importance in determining the thought. In the social and language sciences, knowledge is also studied to measure variables and psychological factors for thought and its interrelationships.

Keywords:- Discourse macrostructure, text network analysis, discourse bias, mental models, topic modelling, text mining, mathematic stats.

I. INTRODUCTION

The process of discussing an issue, expressing one's views and ideas, and persuading others is a mental act of creating knowledge and expressing knowledge, and it is a manifestation of language use. Therefore, modern researchers study ideas in many ways, based on discourse theory and semantics. Since the study of discourse theory and discursive approaches based on linguistic and psychological factors is interesting and modern, researchers are trying to study the interrelationship between statistical methods and language issues. So, this study aims to determine whether a person's thought and knowledge can be determined by mathematical optimization of any decision, according to the use of the language and the semantic structure, and to determine the global meaning and tone of speech at the macrostructure of discourse.

II. LITERATURE REVIEW

A. Language use for communication

Language is the tool for conveying one's thoughts about the role of language. (Wisniewski, 2007) Accordingly, the primary functions of language begin with an understanding of the nature of speech at the level of communication (Bloomfield, 1993) and Finch's (1998) theory (Wisniewski, 2007) considered at the macrostructure.

Function to describe/todorkhoiloh uureg/ - language is used to describe things and actions in the world. (Wisniewski, 2007) People perceive and understand things in their minds from their perspectives, and then the speaker

expresses in words what he wants to say. In other words, when using the function of language to name a phenomenon and object, the speaker begins by describing what has been being said.

Causal function/uchir shaltgaant/ - In theory, the causal function is a means of expressing ideas (Finch 1998, Wisniewski, 2007). Because the human brain processes information, it is very difficult to think without words. When a person thinks about something or a concept, he puts it into words in his mind. For example, if someone thinks about stationery, he thinks of pencils, books, lines, erasers, and pencils. which is expressed in words.

Function of Communication /khariltsaanii uureg/ - This means that most speakers consider language to be a key communication tool. Wanting, begging, informing, commanding, promising, and resisting are all motivated by mutual understanding. (Wisniewski, 2007)

Considering the role of language in the macrostructure, it is responsible for expressing the logic of ideas. These include function of Imagination: Theoretically, it is the process of creating awareness in mental activity, and the principles of understanding and imagining what is happening around people through language. (Finch 1998, Wisniewski, 2007). Textual function is used as a specific language tool and as a template for expressing ideas.

Metafunction: Theoretically, metadiscourse uses the role of language to express discursive approach and author's attitudes play a role in the interaction. (Brown, Yule, 1983). Interactive roles involve social and personal feelings around discourse. (Leech, 1983, хууд. 3) (Wisniewski, 2007) The concept of creating and understanding discourse on the concept of understanding and expressing ideas. (Van Dijk in Wodak & Chilton, 2005, p. 71) The primary goal of the discourse is to use appropriate knowledge to the situation, and as a result the listener understands the situation and achieves mutual results (Van Dijk in Wodak & Chilton, 2005, p. 72) The concept of creating and understanding discourse is interrelated with the ideas and opinions expressed in the context. (Van Dijk in Wodak & Chilton, 2005, p. 71)

Discourse explores a wide range of cognitive-psychological fields, such as relationship formation, text, semantics, and key factors influencing the psychology of the parties involved. (Van Dijk in Wodak & Chilton, 2005, p. 72) From above perspectives, purposeful communication, speech strategies, and speech methods within a given topic can be understood as discourse. (Ibid, p. 5)

III. THEORY OF DISCOURSE STRUCTURE

Discourse is a way of representing and interpreting ideas. The idea of the study of speech and text originated in the 19th century and the first half of the twentieth century (В.Фон Хумболдт, А.А.Потебня, Ф.Боас, В.Я.Пропп, В.Матесниус, С.Балли), and different approaches to speech and text. The idea of studying (or linguistic analysis of text) was developed in the middle of the twentieth century (IA Figurovsky, Z. Harris, K. Pike, P. Hartmann and others). From the 1970s to the 1980s, speech analysis became as important as linguistic research. (В. Дресслер, Т. ван Дайк, Э. Шеглофф, Р. Лонгакр, Т. Гивон, В. Чафе) Maintaining the Integrity of the Specifications.

IV. THEORY AND METHODS OF MODELING KNOWLEDGE

Knowledge is one of the main issues of discourse and includes general and specific, semantic and episodic forms of knowledge of the parties involved in the relationship. Knowledge is a core factor in creating discourse and understanding discourse which is the general "understanding" level of a writer, speaker, or interlocutor, how people discuss the issue from many views, how well it meets the general basic principles, how evaluates the relationship between the real world and the phenomenon, It is the timing of the issue between the times, and how chronology is used, are the most important criteria for determining the direction of an individual's attitude and opinion.

(Б.Дагиймаа) . The understanding is more important than creating discourse.(Нунань, 1993). The general principles and concepts of knowledge creation are similar to the general concepts and principles of knowledge and are defined by theory and models of cognitive psychology, philosophy of knowledge, and general theory of mind.

One of the ways in which knowledge is constructed and analyzed is through the study of discourse and speech. (Alvesson, Karreman, 2000a). The main concept of this field is that language can be used more creatively than patterns from simple to complex, a canonical model, and an approach based on certain constructivist epistemology is used in the study of discourse and speech. (Wood, Kroger, 2000) Also, due to new trends in linguistics, new research styles are being incorporated into social change and new topics and styles are being created, which is a way to study discourse and speech. Exploring the discursive approach and sentiment of discourse based on structure and patterns of models has become a relatively new approach and presents many effective kinds of research. (Fineman, 1996; Mumby, Putnam, 1992).

Text structure theory: Regarding the text structure theory, a theory that explains the logical connection and coherence of meanings in a text-based on the principle that each part of the text and speech has its special function, there are no words or meanings, a hierarchical meaning structure, and a unified introduction and conclusion. Text Rhetorical structure is the sequence of words, reasons, situations, concessions, arguments, explanatory notes,

examples, and conclusions. Parts of the structure have the role of discourse macrostructure. Most relationships are nonlinear and binary. The most obvious part of a macrostructure that comes to mind after reading is the use of terms such as global meaning, macrostructure, and macro propositions in discourse studies. In other words, "memory for the text" is the general concept of meaning that remains in the mind after reading. The "global" meaning of the discourse is explored by modeling the general idea through "topics" and "just". It is a hierarchical structure of meaning that creates the logical sequence of text and speech unity and content. The rhetoric structure is based on the basic principles of integrating the text, the logical connection of meaning, and the relevance of the topic.

Integrating text: semantic and grammatical rules, tools for connecting parts of the text, and their use. Logical relation of meaning: coherence, the logical sequence in structure, optimal wording style, and the correct choice of vocabulary will be considered psychological factors in discourse.

(Reinhart (1980 оны Yuan Wang & Minghe Guo, 2014, pp. 460-465 ишлэл). Organizing any information is a concept of the mind, cognition, and cognitive psychology, and these actions are models for expressing the creation of knowledge. In the macrostructure of the text, the discursive approach is a key factor of speech and even the smallest component of speech creates it. Thus, the discursive approach is a logical and understandable way of speaking, and if the audience understands it well, the purpose of the discourse will be achieved.

The general principles and concepts of knowledge creation are similar to the general concepts and principles of knowledge and are defined by models of cognitive psychology, philosophy of knowledge, and the general theory of mind. The knowledge based on cognitive psychology is mental models in the mind.

A mental model is fact-based knowledge that arises from a basic concept, outcome, or situation. The mental model is a fundamental issue of language, behavior, and cognition, and is a complex set of thoughts and mindsets, such as each individual's perceptions, perspectives, and worldviews. The mental model takes place in the human brain and is an expression of the person's inner knowledge of understanding the world.

(Hall et al. 1994, Swan and Newell 1998, Serman 2000) There is a close relation between language and thought. In the political speech, ideology is discursive; usually in the form of **text** or **speech**. (Dijk, 2006) Therefore, it is the notion that the relationship, the content, and the discourse idea can be interpreted from different views, depending on the method used by the human and the computer.

Mental models, mental maps and mind maps are often used to represent complex ideas and have a positive impact on teaching and learning. Mental maps can be extracted from texts to produce cognitive maps, which can then be

used for sentiment analysis. In the recent years this field has been enhanced with the network metaphor: the nodes are the concepts and the connections are the relations between them. The relations between the concepts can be plotted on the basis of their co-occurrences, which, in turn can facilitate the process of semantic priming — the process by which words tend to be recognized faster when used with the words that have close semantic proximity to them. Text network analysis methods based on this approach have been shown to be successful for better comprehension of texts and topic modeling, (Paranyushkin, 2020).

Topical modeling is a method of calculating the human mind and determining the group of ideas and words. Topical modeling is a key method used in text analysis to determine keyword frequency.

At the written and spoken level, the study of “information distribution and informative properties” by mathematical methods of text analysis and text mining has become an important application in linguistics and communication in determining the mathematical meaning of words and making rational decisions in a short time.

V. ALGORITHMS OF TOPIC MODELING TO REPRESENT MENTAL MODELS

A. IR Models

Text Normalization -Stop words Removal -Boolean Model – TF-IDF – Vector Model – Probabilistic Model – Latent Semantic Indexing Model – Neural Network Model –Retrieval Evaluation – Retrieval Metrics – Precision and Recall – Reference Collection –Userbased Evaluation.

B. Topic modeling to 'tf-idf' algorithm

Text mining algorithms can model models of words, terms, phrases, complete sentences, and complexities. The modern text mining approach to textual research is becoming an innovative style of research in this area. Text mining is used to analyze text clusters, text macro structures, sentiment analysis, and conversion of spoken text into numeric values. It is a “mathematical method and programming language” interface model for any presentation, speech, text, or source text. It is an important way to process and analyze information across sectors. ‘Text mining = Methods for extracting useful information from large and often unstructured collections of texts. It is also closely related to “information retrieval.” In this context, keywords that carry information about the contents of a document are called terms. A list of all the terms in a document is called an index. For each term, a list of all the documents that contain that particular term is called an inverted index, a matrix normally represents the frequency of occurrence of term in document.

(Saito, 2013)

C. latent semantic analysis or LSA, pLSA algorithm

latent dirichlet allocation or LDA, relational topic models, word2vec algorithm, and its extension lda2vec. These methods are based on retrieving the topics from text by identifying the clusters of co-occurrent words within them, based on the bag-of-words and skip-gram models. This data can then be used to classify similar documents, improve text indexing and retrieval methods, and to identify evolution of certain topics over a period of time within a specific text corpus.

LDA provides the ability to get a better understanding of the text network topology, which can yield many other insights about the discourse structure in LDA or word2vec: how connected a discourse is, whether it’s biased towards a certain set of concepts, how dominant a specific topic is towards the whole discourse, whether there are structural gaps that may indicate a potential for new ideas. (Paranyushkin, 2020)

D. Text-to-Network Conversion

The text is then converted into a directed network graph. The normalized words (lemmas) are the nodes in the network graph and their co-occurrences are the edges.

a) Extracting Most Influential Keywords

Using Betweenness Centrality, apply a ranking algorithm to identify the nodes (the words) with the highest betweenness centrality – these are the nodes that appear most often on the shortest paths between any two. Two consecutive scans are performed. The first scan creates the connections (graph edges) between the lemmas (nodes) that appear next to each other (bigrams) with the edge weight value (Paranyushkin, 2020).

Those words are central for meaning circulation and can be seen as the meaning junctions within the discourse — they are shown bigger on the graph (Paranyushkin, 2020).

b) Research case:

Speech by Ts. Elbegdorj president of Mongolia at the general debate of the 71th session of the united nations general assembly on “The sustainable development goals: a universal push to transform our world” (Elbegdorj, 2006)

node name	degree	frequency	betweenness	topic	diversivity
sum total	1752	688	2.770113	n/a	3787.6
sum / 150 nodes	11.68	4.59	0.018467	n/a	25.25
mongolia	57	25	0.357274	2	142.9
people	42	16	0.174323	2	109
development	38	17	0.170083	3	100
human right	34	19	0.147893	7	77.8
government	22	9	0.1397	3	155.2
world	32	14	0.12048	2	86.1
sustainable	17	8	0.092962	4	116.2
international	27	9	0.07734	6	85.9
law	24	11	0.075753	4	68.9
peace	22	10	0.067855	6	67.9
sustainable	22	9	0.060206	3	66.9

Table 1: Discourse semantic network table

Data table of Semantic macrostructure of “The sustainable development goals: a universal push to transform our world” (Elbegdorj, 2006)

Frequency: values with semantic significance in the discourse structure: 1752 /150 nodes,

The most Significant values: between topics

“MONGOLIA” 57/25/0.357/2

“PEOPLE” 42/16/0.174/2

“DEVELOPMENT”38/17/0.17

The network's structure diversity is higher which is its mind-viral immunity — that is, such a network will be more resilient and adaptive than a less diverse one. In case of a discourse network, high mind-viral immunity means that the text proposes multiple points of view and propagates its influence using both highly influential concepts and smaller, secondary topics.

Modularity: 0.53 (medium)

Influence Distribution 80%

(the measure of influence distribution (the entropy of the top nodes' distribution among the top clusters),

as well as the the percentage of nodes in the top community)

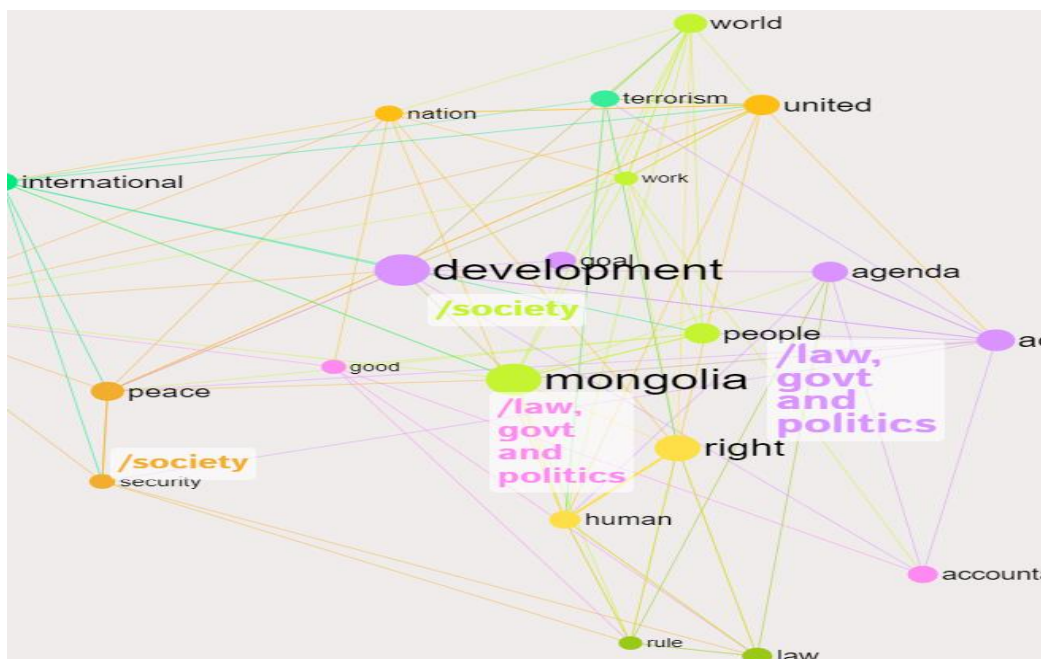


Fig.1: Discourse semantic network and topics

The topics are the nodes (words) that tend to co-occur together in the same context (next to each other). Detection algorithm: a combination of clustering and graph community detection algorithm to identify the groups of nodes are more densely connected together than with the rest of the network.

They are aligned closer to each other on the graph are given a distinct color.

The top prominent nodes that have significantly higher influence than the rest.

The most significant nodes are either the ones with the highest betweenness centrality (current setting) — appearing most often on the shortest path between any two randomly chosen nodes are ‘MONGOLIA’, ‘PEOPLE’, ‘DEVELOPMENT’

The higher is the diversity, the more distinct communities (topics) there are in this network, the more likely it will be pluralist.

Topical Cluster	Percentage of Nodes	Category
1	19%	/society
2	19%	/law, govt and politics
3	15%	/government
4	14%	/computer security
5	10%	/society
6	7%	/law, govt and politics
7	5%	/legislation
8	5%	culture environmental nomadic
9	3%	climate change

Table II. Shows the percentage of Topics of Fig.1

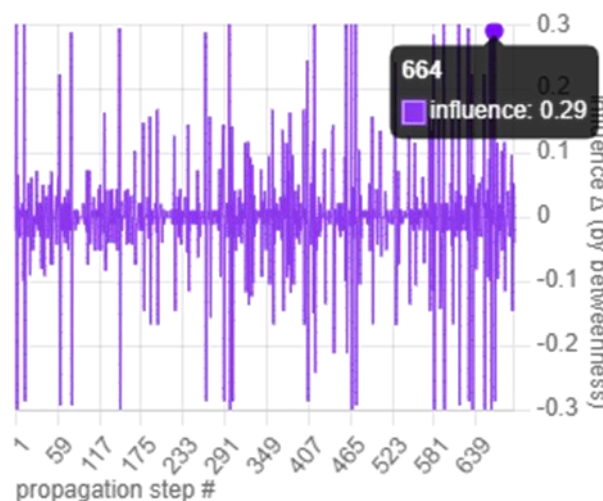


Fig. 2: Influence betweenness topics

The fig.2 Shows how influence propagates through the network.

X-axis: lemma to lemma step (narrative chronology).

Y-axis: change of influence.

The more even and rhythmical this propagation is, the stronger is the central idea or agenda Propagation dynamics: irregular variability | alpha exponent: 0.60 | high (based on Detrended Fluctuation Analysis of influence)

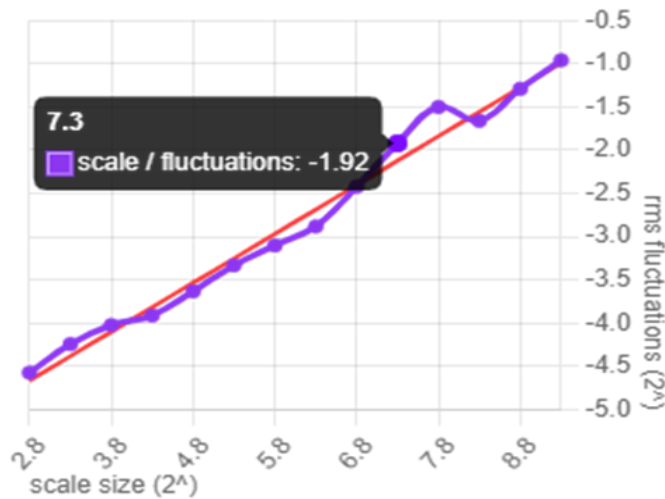


Fig. 2: Influence fluctuation scale between topics

Propagation dynamics: Irregular variability | alpha exponent: 0.60 | high
(based on Detrended Fluctuation Analysis of influence)

The more variability can be seen in the propagation profile, the less is the reliance on the main concepts (agenda), the stronger is the role of secondary topical clusters in the narrative.

We plot the narrative as a time series of influence (using the words' betweenness score). We then apply detrended fluctuation analysis to identify fractality of this time series. Plotting the log₂ scales (x) to the log₂ of accumulated fluctuations (y).

ks: 1.42, d: 0.37 > cr: 0.35 | power law not identified (based on kolmogorov-smirnov test)

If a power-law is identified, the nodes have preferential attachment (e.g. 20% of nodes tend to get 80% of connections), and the network may be scale-free, which may indicate that it's more resilient and adaptive. Absence of power law may indicate a more equalized distribution of influence.

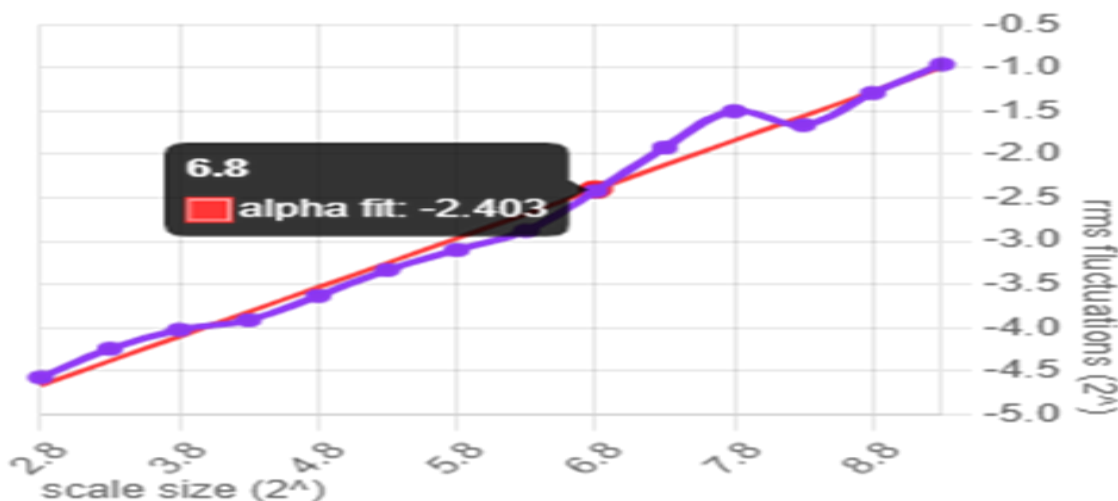


Fig. 3: Influence fluctuation Alpha fit between topics

Using the alpha exponent of the fit (which is closely related to Hurst exponent)),

- relation: uniform (pulsating | alpha <= 0.65),
- variable (stationary, has long-term correlations | 0.65 < alpha <= 0.85), fractal (adaptive | 0.85 < alpha < 1.15), and complex (non-stationary | alpha >= 1.15).
- "fractal" (near-critical state): maximal diversity, adaptivity, and plurality, the narrative
- "uniform" (For fiction, essays, and some forms of poetry)
- "variable + stationary" (Informative)
- "complex" (state is an indicator or shifting its state)

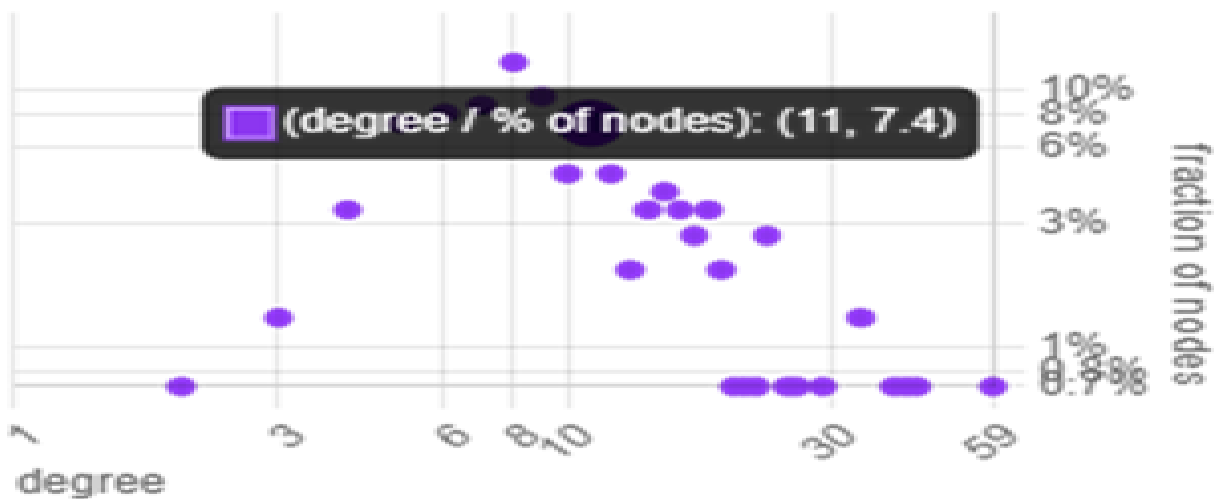


Fig. 4: Fraction of nodes

ks: 1.42, d: 0.37 > cr: 0.35 | power law not identified (based on kolmogorov-smirnov test)

If a power-law is identified, the nodes have preferential attachment (e.g. 20% of nodes tend to get 80% of connections), and the network may be scale-free, which may indicate that it's more resilient and adaptive. Absence of power law may indicate a more equalized distribution of influence.

In combination with the measure of influence distribution (the entropy of the top nodes' distribution among the top clusters), as well as the the percentage of nodes in the top community is below:

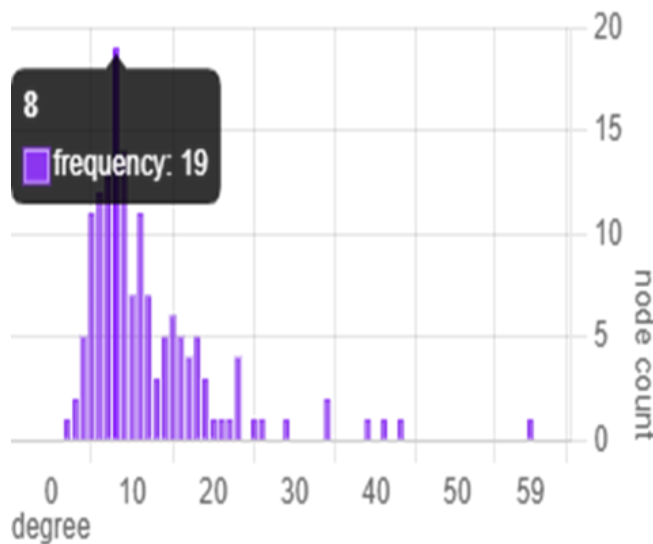


Fig. 5: Scaled nodes network

ks: 1.42, d: 0.37 > cr: 0.35 | power law not identified (based on kolmogorov-smirnov test)

(network has scale-free / small-world (long-tail power law distribution) or random (normal, bell-shaped distribution) network properties.

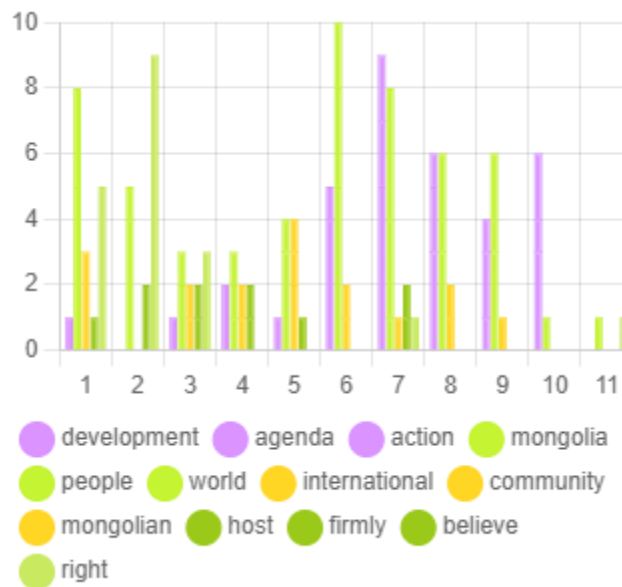


Fig. 5: Evolution of topics in discourse

The chart shows how the main topics and the most influential keywords evolved over time.
 X-axis: time period (split into 10% blocks).
 Y-axis: cumulative frequency of occurrence.

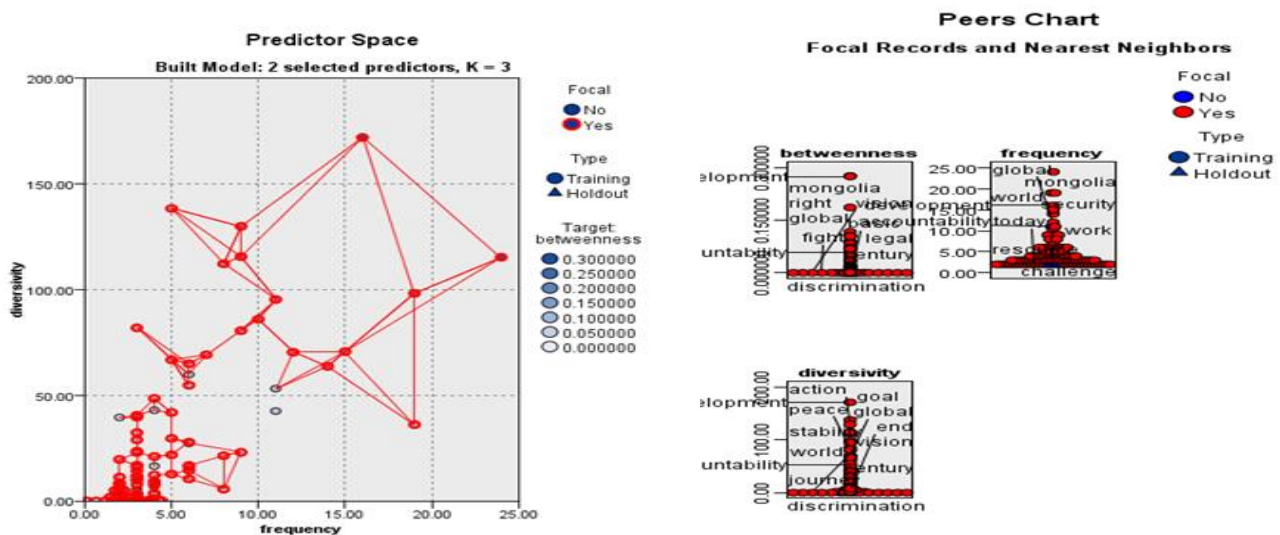


Fig. 6: K-NN scaling, The nearest semantic neighbors of predicted space

VI. METHODS FOR NETWORK ANALYSIS

The Method is an open-source software written in JavaScript using Sigma.js, Cytoscape and Graphology libraries in the front-end and java-based Neo4J graph database for the backend, and SPSS: K-NN.

VII. SUMMARY

Regarding the text structure theory, a theory that explains the logical connection and coherence of meanings in a text-based on the principle that each part of the text and speech has its special function, there are no words or meanings, a hierarchical meaning structure, and a unified introduction and conclusion.

Integrating text: semantic and grammatical rules, tools for connecting parts of the text, and their use. Logical relation of meaning: coherence, the logical sequence in structure, optimal wording style, and the correct choice of vocabulary will be considered psychological factors in discourse.

A mental model: fact-based knowledge arises from a basic concept, which is a fundamental issue of language, behavior, and cognition, and is a complex set of thoughts and mindsets, (Б.Дагиймаа). **Mental models**, mental maps and mind maps are often used to represent complex ideas and have a positive impact on teaching and learning, can be extracted from texts to produce cognitive maps, which can then be used for sentiment analysis. (Paranyushkin, 2020).

Topical modeling: Topical modeling is a key method used in text analysis to determine keyword frequency.

VIII. RESULT OF NETWORK ANALYSIS

The network's structure diversity is higher, which is its mind-viral immunity

Modularity: 0.53
Influence distribution: 80%

The topics are the nodes (words) that tend to co-occur together in the same context (next to each other). Detection algorithm: a combination of clustering and graph community detection algorithm to identify the groups of nodes are more densely connected together than with the rest of the network.

The most significant nodes are either the ones with the highest betweenness centrality (current setting) — appearing most often on the shortest path between any two randomly chosen nodes are 'MONGOLIA', 'PEOPLE', 'DEVELOPMENT'

The higher is the diversity, the more distinct communities (topics) there are in this network, the more likely it will be pluralist.

X-axis: lemma to lemma step (narrative chronology).
Y-axis: change of influence.

Propagation dynamics: irregular variability | alpha exponent: 0.60 | high (based on Detrended Fluctuation Analysis of influence) plotting the log₂ scales (x) to the log₂ of accumulated fluctuations (y).

ks: 1.42, d: 0.37 > cr: 0.35 | power law not identified (based on kolmogorov-smirnov test)
relation: uniform (pulsating | alpha <= 0.65),

variable (stationary, has long-term correlations | 0.65 < alpha <= 0.85), fractal (adaptive | 0.85 < alpha < 1.15), and complex (non-stationary | alpha >= 1.15). "fractal" (near-critical state): maximal diversity, adaptivity, and plurality, the narrative.

This discourse's structure is balanced representation of several perspectives and its immunity level is high.

It can be made more focused if develop the ideas around the main keywords — MONGOLIA, PEOPLE, DEVELOPMENT, RIGHT — or by identifying the structural gap between the less represented topics

X-axis: time period (split into 10% blocks).
Y-axis: cumulative frequency of occurrence.

The nearest semantic neighbors of predicted space, identified by the factor and variables with K-NN SPSS

ACKNOWLEDGMENTS

This research work was supported by text visualization and topic modeling algorithms on nodus source and SPSS platform. We also want to thank for providing access to their resources for research purposes. As well, thanks to speech by Ts. Elbegdorj president of Mongolia at the general debate of the 71th session of the united nations general assembly on "The sustainable development goals: a universal push to transform our world" (Elbegdorj, 2006)

REFERENCES

- [1.] B.Dagiimaa, & O.Munkhchimeg. (2022). Text Mining Applications in Discourse Analysis and Modern Trends (Big Data Methods on Political Speeches)
- [2.] Bloomfield. (1998). *The MIT Encyclopedia of Cognitive Science*. (F. C. Robert A. Wilson, Хял.) MIT PressEditors.
- [3.] F.Ungerer, & H.J. Schmid. (2012). *An introduction to Cognitive Linguistics*. (E. P. M.A.K. Halliday., Хял.) Foreign Language Teaching and Research Press.
- [4.] F.Ungerer, H. S. (2012). *An introduction to Cognitive Linguistics*. (E. P. M.A.K. Halliday.) . Foreign Language Teaching and Research Press.
- [5.] Fairclough. (2017). Critical Approach In Social Research. *The Online Journal of Communication and Media*.
<https://www.tojned.net/journals/tojcam/articles/v03i04/v03i04-02.pdf-ээс> Гаргасан
- [6.] Van Dijk in Wodak & Chilton, 2005, p. 71-72. *Critical Discourse Analysis*.
- [7.] Б.Дагиймаа. (2020). *Дискурсын онол арга зүйн асуудлууд: ярилцлагын дискурсын жишээгээр* (Б. 4/2020). Улаанбаатар хот: Европ судлал, эрдэм шинжилгээний сэтгүүл.
- [8.] Staggers, Nancy; Norcio, A.F. (1993). "Mental models: concepts for human-computer interaction research". *International Journal of Man-Machine Studies*.
- [9.] M. Young, *The Technical Writer's Handbook*. Mill Valley, CA: University Science, 1989.
- [10.] Paranyushkin, D. (2020). *InfraNodus: Generating Insight Using Text Network*. Paris, France.
- [11.] Elbegdorj, T. (2006). "THE SUSTAINABLE DEVELOPMENT GOALS: Ulaanbaatar, Mongolia: https://www.un.int/mongolia/statements_speeches/.