Rahva Teknik ve Sosyal Araştırmalar Dergisi
Cilt: 1, Sayı: 2, 2021, Sayfa: 84 - 91

Rahva Journal of Technical and Social Studies
Vol: 1, Issue: 2, 2021, Page: 84 - 91

# Centrality with Entropy in Hypergraphs*

İhsan TUĞAL[1, a, ✉], Zeydin PALA[1, b]

[1]Muş Alparslan University, Faculty of Engineering and Architecture, Software Engineering, Muş, Turkey
ORCID: [a]0000-0003-1898-9438; [b]0000-0002-2642-7788

[✉]Sorumlu Yazar: i.tugal@alparslan.edu.tr

**Abstract:** Entropy is used to measure uncertainty in complex systems. Hypergraphs provide structure for mathematically modeling real-world data. In this study, analyzes were made using entropy on the data in the hypergraph structure. The entropies of the nodes and hyperedges were calculated using the node degree and hyper edge degree. Their activities were found according to these values. The applicability of this method in weighted or unweighted relational structures was demonstrated through examples. In institutions with multiple departments and employees, the results obtained with the proposed method can be used to support decision-making processes.

**Keywords:** *Entropy, Hypergraphs, Decision Support Systems, Centrality*

# Entropi ile Hiper Çizgelerde Merkezilik

**Özet:** Entropi kompleks sistemlerde belirsizliği ölçmek için kullanılabilir. Hiper çizgeler gerçek dünyaya uygun verileri matematiksel olarak modellemek için yapı sunar. Bu çalışmada hiper çizge yapısındaki veriler üzerinde entropi kullanılarak analizler yapıldı. Düğüm derecesi ve hiper kenar derecesi kullanılarak düğümlerin ve hiper kenarların entropileri hesaplandı. Bu değerlere göre etkinlikleri bulundu. Ağırlıklı veya ağırlıksız ilişkisel yapılarda bu yöntemin uygulanabilirliği örnekler üzerinden gösterildi. Birden çok birimi ve çalışanı olan kurumlarda karar verme süreçlerinde destek amaçlı önerilen yöntemle elde edilen sonuçlar kullanılabilir.

**Anahtar Sözcükler:** *Entropi, Hiper Çizgeler, Karar Destek Sistemleri, Merkezilik*

## 1. Introduction

Graphs are one of the main approaches used especially for pattern recognition and machine learning tasks (Martino & Rizzi, 2020). Although hypergraphs are ubiquitous, their concepts are less well known than graphs, and they are often used unspecified (Klamt et al., 2009). It is also commonly used to reduce data in hypergraph structure to normal graphs (Tuğal et al., 2013). More use of hypergraphs and their mathematical analysis will provide more accurate analyzes without loss of information.

In this study, influential nodes and hyperedges were determined with the proposed method using entropy, which is an uncertainty measure on hypergraphs. The proposed method was applied to a synthetic data. These and similar methods can be used to detect patterns, find the key points of a system, analyze relational structures, and support multidimensional decision-making processes.

In the second part of our study, studies in the literature are mentioned. Hypergraphs and entropy were explained in the third and fourth chapters. In the fifth section, the proposed method was applied to the synthetic data. In the last section, the results of the study were mentioned.

## 2. Related Work

Bonacich et al. used hypergraphs in their network analysis of the attacks of the indigenous inhabitants of the Caribbean Island on Spanish settlements (Bonacich et al., 2004).

*İhsan TUĞAL, Zeydin PALA, Identifying Influential Hyper Edges and Nodes with Entropy in Hypergraphs, 1st International Rahva Technical And Social Researches Congress, 4-5 December 2021, ISBN: 978-605-74216-1-6, pp.: 236-237*

For heterogeneous hypergraphs, a graph neural network-based representation learning framework is proposed, which is an extension of traditional graphs that can well characterize multiple non-binary relationships (Sun et al., 2021).

Graph-theoretic techniques for the holographic entropy cone were generalized to study hypergraphs and similarly defined entropy cones. This allows developing a framework for calculating entropies efficiently and proving the inequalities provided by hypergraphs (Bao, Cheng et al., 2020).

Hu et al. have described the mathematical properties of the hypergraph by presenting a definition similar to the entropy calculation used in this study. The extremality of entropy of graphs according to the degrees of uniform hyper graphs has been examined (Hu et al., 2019).

In the study, a new entropy concept was developed for uniform hypergraphs based on tensor theory. Results were constructed on the lower and upper bounds of entropy, and it is shown to be a measure of regularity for uniform hypergraphs with two simulated examples, based on node degrees, path lengths, clustering coefficients, and negligible symmetry (Chen & Rajapakse, 2020).

In another work, a new definition of entropy for hypergraphs was introduced. The fine structure of these graphs has been taken into account by considering partial hypergraphs that give an entropy vector. The properties of the proposed definitions for hypergraphs have been analyzed (Bloch & Bretto, 2019).

## 3. Hypergraphs

A graph is an example of a structure where a series of nodes are connected by edges. It is an organized structure. Graphs can be expressed with an adjacency matrix. The full set of all binary connections defines the topology of the graph by providing a complete map of all relationships between nodes and edges (Sporns, 2018). Such topological and semantic data structures are widely used to model systems such as telecommunications, sociology, biological and social networks (Tuğal & Karcı, 2020).

Although there are many problems that graphs can solve, the disadvantage is that they only take into account pairwise relations. By definition, an edge can only connect two nodes. This limits the modeling power of graphs. It causes incomplete expression of data. An incomplete definition emerges. Hypergraphs overcome these limitations by allowing more than two nodes to be connected simultaneously with the hyper edges solution. Thanks to hyper edges, we are able to express the multidirectional relationships between nodes. Better modeling capabilities of versatile relationships have been demonstrated in fields such as biology (eg. protein-protein interaction networks) and social networks (eg. collaboration networks) (Klamt et al., 2009). It plays an important role to analyze a set of data using techniques derived from topology and mathematics. In addition, the computational complexity, space size and cost are lower (Wolf et al., 2016).

Hypergraphs are generalizations of graphs whose edges contain more than one node and thus represent k-way relationships. Therefore, hypergraphs represent many samples that can be taken from natural life as a dataset. Hyper graph structured data is ubiquitous. Set-valued, tabular, or bipartite data can naturally be represented by hypergraphs. It has features related to a number of mathematical structures that are important in data science. The uncertainties special to these complex structures can be resolved if strong mathematical methods are used. (Aksoy et al., 2020).

As shown in Figure 1 above, a hypergraph can be defined as $H=(V,E)$. V stands for the set of nodes and E stands for the set of hyper edges. $n=|V|$ is the number of nodes of the hypergraph and $m=|E|$ is the number of hyper edges. For each $v \in V$, the set of hyper edges containing v node is denoted by $E(v) \subseteq E$. The node degree can be shown with $d(v)=|E(v)|$.

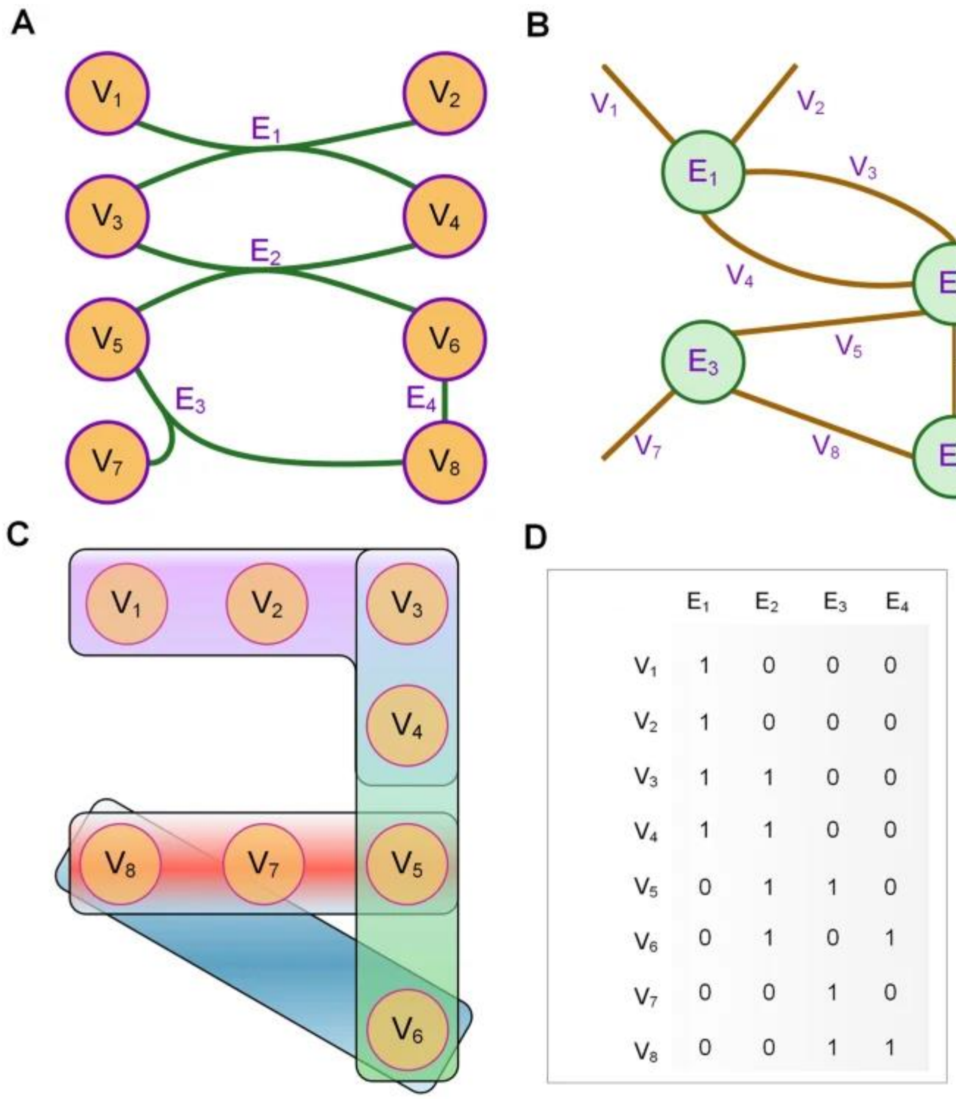


**Figure 1.** Representation of hypergraphs (Shen et al., 2018).

## 4. Entropy

First entropy has been used for the measurement of uncertainties in thermodynamics (Boltzmann, 1964) and then communication theory (Shannon, 1951; Shannon, 1948). Since it is based on probabilistic computation, it has been widely used in other disciplines (Bromiley et al., 2004). Many different definitions of entropy can be made (Deng, 2016; Karci, 2016; Karcı, 2018; Rényi, 1961; Tsallis, 2013). It is used in the analysis of structural data in decision making processes. Entropy is used to measure structural complexity in graphs and to interpret relational uncertainty ( Tuğal & Karcı, 2019). In this study, entropy was used for hypergraph analysis.

If the probability of an event occurring is represented by p, the total entropy value is calculated as in equation 1 by using the probability values of n events.

$$\boldsymbol{I}(\boldsymbol{p_1}, \boldsymbol{p_2}, \ldots\ldots, \boldsymbol{p_n}) = -\sum_{i=1}^{n} \boldsymbol{p_i} \boldsymbol{log} \boldsymbol{p_i} \tag{1}$$

In this study, while calculating the total entropy value of each node or hyper edge, the values in the I_ij matrix were used, which represents the hypergraph. e denotes for hyper edges, v denotes for nodes. d(v) denotes the node degree and d(e) denotes the hyper edge degree. In weighted structures, the degree is determined by calculating the weights. More precise measurements are obtained.

**Table 1** Incidence matrix and degrees

| | $e_1$ | $e_2$ | $e_3$ | $d(v)$ |
|---|---|---|---|---|
| $v_1$ | 1 | 1 | 0 | 2 |
| $v_2$ | 0 | 1 | 0 | 1 |
| $v_3$ | 1 | 1 | 1 | 3 |
| $v_4$ | 1 | 0 | 0 | 1 |
| $v_5$ | 0 | 0 | 1 | 1 |
| $d(e)$ | 3 | 3 | 2 | |

Using the values in Table 1, the entropy values for both nodes and hyper edges can be calculated with the following equations 2 and 3. These values indicate the effective of nodes or hyperedges in the relational system.

$$I(v_1) = -\sum_{j=1}^{n} \frac{|I_{1j}|}{d(v_1)} log \frac{|I_{1j}|}{d(v_1)} = -(\frac{1}{2} log \frac{1}{2} + \frac{1}{2} log \frac{1}{2}) = 1 \quad (2)$$

$$I(e_1) = -\sum_{i=1}^{m} \frac{I_{i1}|}{d(e_1)} log \frac{|I_{i1}|}{d(e_1)} = -(\frac{1}{3} log \frac{1}{3} + \frac{1}{3} log \frac{1}{3} + \frac{1}{3} log \frac{1}{3}) = 1.585 \quad (3)$$

By identifying effective nodes/edges, it can be understood which patterns to focus on and which ones to look at in detail.

## 5. Experiments

Our synthetic dataset aims to analyze the staffs over the departments they work in. It is a data consisting of 5 departments and 10 personnel working in these departments. It is stated that these personnel contribute to which departments and to what size. This data, shown in the hypergraph structure, was analyzed using entropy.

**Table 2** Department and staff weighted incidence matrix

| | | | **DEPARTMENT** | | | | |
|---|---|---|---|---|---|---|---|
| | | | **Network** | **System** | **Software** | **Technical** | **Administrative** |
| | | | e0 | e1 | e2 | e3 | e4 |
| **STAFF** | **A** | v0 | 0,1 | 0,3 | 0 | 0,6 | 0 |
| | **B** | v1 | 0 | 1 | 0 | 0 | 0 |
| | **C** | v2 | 0 | 0 | 1 | 0 | 0 |
| | **D** | v3 | 1 | 0 | 0 | 0 | 0 |
| | **E** | v4 | 0 | 0 | 1 | 0 | 0 |
| | **F** | v5 | 0,2 | 0,2 | 0,4 | 0,1 | 0,1 |
| | **G** | v6 | 0 | 0,5 | 0,5 | 0 | 0 |
| | **H** | v7 | 0,5 | 0,5 | 0 | 0 | 0 |
| | **I** | v8 | 0,5 | 0,5 | 0 | 0 | 0 |
| | **K** | v9 | 0 | 0 | 0 | 0 | 1 |

The aim is to determine the weight of the personnel or unit in the IT department. Of course, this weight measurement is based on the available data. To examine the relational structure of the staff or department. This system can be made more decisive by adding other parameters, such as which programming languages they know or which certificates they have, and criteria that measure their contribution to the projects. With this kind of information obtained, the aim is to make managerial analyzes, to organize the distribution of tasks, to understand the impact of the staff or the importance of the department, etc.

For this purpose, entropy was used in our study. Entropy is a measure of uncertainty. It can be used in decision making processes, machine learning applications and etc. (Aggarwal, 2021; Hark & Karcı, 2020).

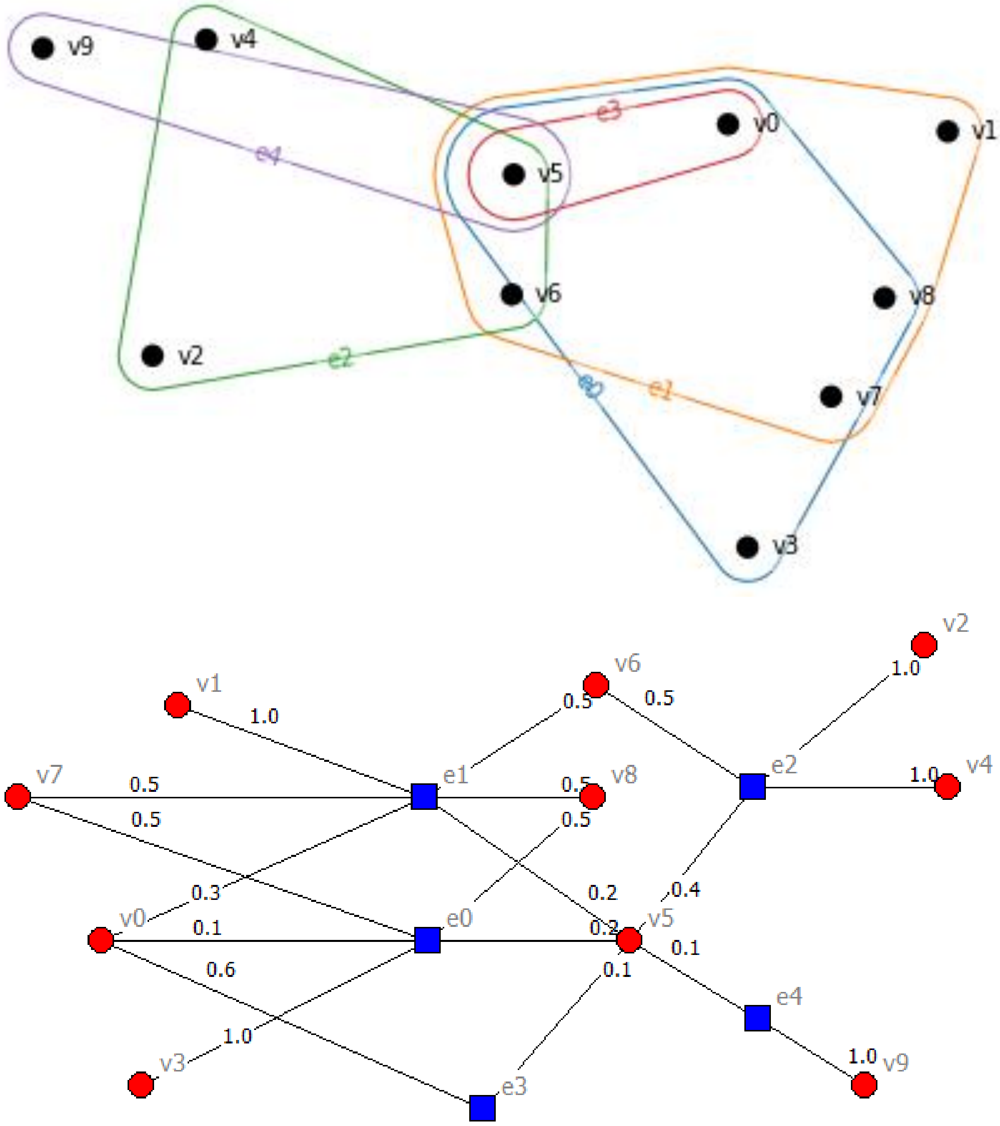


**Figure 2.** Hypergraph (Borgatti et al., 2002; Praggastis et al., 2019)

When the values obtained in entropy calculations are examined, it is seen that person F is the most weighted personnel due to his contribution to the department. The second most weighted personnel was Staff A. Since people B, C, D, E, K contribute to only one department and they are not very much in this relational structure, their entropy values are 0. Persons G, H and I took the value 1 because they are associated with two departments.

When we look at the departments, the entropy value of the system department is the highest because it receives support from more staff. The network department came in second as it received support from 5 staff. We understand that these units need more personnel support with different skills.

As seen in Table 3, these entropy values produced results by taking into account the relationship density first and the relationship weight in the second step. The effect of hyper edges and nodes in hypergraphs could be measured by entropy.

**Table 3** Department and staff entropy values

| Staff | | Entropy |
|---|---|---|
| **A** | v0 | 1,29546 |
| **B** | v1 | 0 |
| **C** | v2 | 0 |
| **D** | v3 | 0 |
| **E** | v4 | 0 |
| **F** | v5 | 2,12193 |
| **G** | v6 | 1 |
| **H** | v7 | 1 |
| **I** | v8 | 1 |
| **K** | v9 | 0 |

| Department | | Entropy |
|---|---|---|
| **Network** | e0 | 1,98275 |
| **System** | e1 | 2,41345 |
| **Software** | e2 | 1,8908 |
| **Technical** | e3 | 0,591673 |
| **Administrative** | e4 | 0,439497 |

## 6. Conclusion

In this study, proposed method performed the analysis on two-dimensional data modeled as a hypergraph. It has been shown that we can detect influential nodes or hyper edges in hypergraphs by entropy. Hypergraphs provide a structure that is more appropriate to real-world data than graphs. We have seen more clearly with the application that the use of hypergraphs to model the data reduces the disadvantages such as loss of information. The method can also be applied to weighted hypernetworks. It was observed that the entropy values were affected by the density of the relations and the weight of the links, respectively. Therefore, making analyzes with hypergraphs provides more accurate results. In addition, time and space complexity are lower. Entropy gives us results in centrality measurements over uncertainty. It shows the relationship between uncertainty and central nodes in a structure. Centrality measurements with entropy always produce accurate results in any structure. Our next goal is to further develop this method and apply it to biological data. It is to contribute to the more widespread use of hypergraphs.

### Acknowledgements

The summary of the study was presented at 1st Rahva Technical and Social Researches Congress 2021.

### Financial Support

This research did not receive any grants from any funding institution/industry.

### Conflict of Interest

The authors declared that there is no conflict of interest.

### References

Aggarwal, M. (2021). Attitude-based entropy function and applications in decision-making. *Engineering Applications of Artificial Intelligence, 104,* 104290. https://doi.org/10.1016/j.engappai.2021.104290

Aksoy, S. G., Joslyn, C., Ortiz Marrero, C., Praggastis, B., & Purvine, E. (2020). Hypernetwork science via high-order hypergraph walks. *EPJ Data Science, 9*(1), 16. https://doi.org/10.1140/epjds/s13688-020-00231-0

Bao, N., Cheng, N., Hernández-Cuenca, S., & Su, V. P. (2020). The quantum entropy cone of hypergraphs. *SciPost Physics*, *9*(5), 067. https://doi.org/10.21468/SciPostPhys.9.5.067

Bloch, I., & Bretto, A. (2019). A New Entropy for Hypergraphs (pp. 143–154). https://doi.org/10.1007/978-3-030-14085-4_12

Boltzmann, L. (1964). *Lectures on Gas Theory*. Berkeley: University of California Press.

Bonacich, P., Cody Holdren, A., & Johnston, M. (2004). Hyper-edges and multidimensional centrality. *Social Networks*, *26*(3), 189–203. https://doi.org/10.1016/j.socnet.2004.01.001

Borgatti, S. P., Everett, M. G., & Freeman, L. C. (2002). UCINET 6 For Windows: Software for Social Network Analysis, Analytic Technologies, Harvard, MA. *Analytic Technologies*.

Bromiley, P. A., Thacker, N. A., & Bouhova-Thacker, E. (2004). Shannon entropy, Renyi entropy, and information. Retrieved December 14, 2018, from http://www.tina-vision.net/docs/memos/2004-004.pdf

Chen, C., & Rajapakse, I. (2020). Tensor Entropy for Uniform Hypergraphs. *IEEE Transactions on Network Science and Engineering*, *7*(4), 2889–2900. https://doi.org/10.1109/TNSE.2020.3002963

Deng, Y. (2016). Deng entropy. *Chaos, Solitons and Fractals*, *91*, 549–553. https://doi.org/10.1016/j.chaos.2016.07.014

Hark, C., & Karcı, A. (2020). Karcı summarization: A simple and effective approach for automatic text summarization using Karcı entropy. *Information Processing & Management*, *57*(3), 102187. https://doi.org/10.1016/j.ipm.2019.102187

Hu, D., Li, X. L., Liu, X. G., & Zhang, S. G. (2019). Extremality of Graph Entropy Based on Degrees of Uniform Hypergraphs with Few Edges. *Acta Mathematica Sinica, English Series*, *35*(7), 1238–1250. https://doi.org/10.1007/s10114-019-8093-2

Karci, A. (2016). Fractional order entropy: New perspectives. *Optik*, *127*(20), 9172–9177. https://doi.org/10.1016/j.ijleo.2016.06.119

Karcı, A. (2018). Notes on the published article “Fractional order entropy: New perspectives.” *Optik - International Journal for Light and Electron Optics*, *171*, 107–108.

Klamt, S., Haus, U.-U., & Theis, F. (2009). Hypergraphs and Cellular Networks. *PLoS Computational Biology*, *5*(5), e1000385. https://doi.org/10.1371/journal.pcbi.1000385

Martino, A., & Rizzi, A. (2020). (Hyper)graph Kernels over Simplicial Complexes. *Entropy*, *22*(10), 1155. https://doi.org/10.3390/e22101155

Praggastis, B., Arendt, D., Joslyn, C., Purvine, E., Aksoy, S., & Monson, K. (2019). HyperNetX. Retrieved from https://github.com/pnnl/HyperNetX

Rényi, A. (1961). On measures of entropy and information. In *Fourth Berkeley Symposium on Mathematical Statistics and Probability* (p. 547). https://doi.org/10.1021/jp106846b

Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Shannon, C. E. (1951). Prediction and Entropy of Printed English. *Bell System Technical Journal*, 30(1), 50–64. https://doi.org/10.1002/j.1538-7305.1951.tb01366.x

Shen, T., Zhang, Z., Chen, Z., Gu, D., Liang, S., Xu, Y., … Xie, X. (2018). A genome-scale metabolic network alignment method within a hypergraph-based framework using a rotational tensor-vector product. *Scientific Reports, 8*(1), 16376. https://doi.org/10.1038/s41598-018-34692-1

Sporns, O. (2018). Graph theory methods: applications in brain networks. *Dialogues in Clinical Neuroscience, 20*(2), 111–120. https://doi.org/10.31887/DCNS.2018.20.2/osporns

Sun, X., Yin, H., Liu, B., Chen, H., Cao, J., Shao, Y., & Viet Hung, N. Q. (2021). Heterogeneous Hypergraph Embedding for Graph Classification. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining* (pp. 725–733). New York, NY, USA: ACM. https://doi.org/10.1145/3437963.3441835

Tsallis, C. (2013). Entropy. In *Computational Complexity: Theory, Techniques, and Applications* (Vol. 9781461418, pp. 940–964). https://doi.org/10.1007/978-1-4614-1800-9_61

Tuğal, İ., & Karcı, A. (2019). Comparisons of Karcı and Shannon entropies and their effects on centrality of social networks. *Physica A: Statistical Mechanics and Its Applications*, 523, 352–363. https://doi.org/10.1016/j.physa.2019.02.026

Tuğal, İ., & Karcı, A. (2020). Determination of Influential Countries by Cultural and Geographical Parameters. *Anemon Muş Alparslan Üniversitesi Sosyal Bilimler Dergisi*. https://doi.org/10.18506/anemon.563211

Tuğal, İ., Kaya, M., & Tuncer, T. (2013). Link prediction in disease and drug networks. In *6-th International Conference of Advanced Computer Systems and Networks: Design and Application (ACSN 2013)* (pp. 46–50).

Wolf, M. M., Klinvex, A. M., & Dunlavy, D. M. (2016). Advantages to modeling relational data using hypergraphs versus graphs. In *2016 IEEE High Performance Extreme Computing Conference (HPEC)* (pp. 1–7). IEEE. https://doi.org/10.1109/HPEC.2016.7761624