INFORMATION THEORY OF CARTOGRAPHY: A FRAMEWORK FOR ENTROPY-BASED CARTOGRAPHIC COMMUNICATION THEORY

Zhilin LI [1, 2]

[1] State-Province Joint Engineering Laboratory of Spatial Information Technology for High-speed Railway Safety, Southwest Jiaotong University, Chengdu, China - dean.ge@swjtu.edu.cn

[2] Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong SAR – lszlli@polyu.edu.hk

Commission VI, WG VI/1

KEY WORDS: Information Theory of Cartography; Cartographic communication, Cartographic Information Theory; Boltzmann entropy; Generalized Shannon Entropy.

ABSTRACT:

Map is an effective communication means. It carries and transmits spatial information about spatial objects and phenomena, from map makers to map users. Therefore, cartography can be regarded as a communication system. Efforts has been made on the application of Shannon Information theory developed in digital communication to cartography to establish an information theory of cartography, or simply cartographic information theory (or map information theory). There was a boom during the period from later 1960s to early 1980s. Since later 1980s, researcher have almost given up the dream of establishing the information theory of cartography because they met a bottleneck problem. That is, Shannon entropy is only able to characterize the statistical information of map symbols but not capable of characterizing the spatial configuration (patterns) of map symbols. Fortunately, break-through has been made, i.e. the building of entropy models for metric and thematic information as well as a feasible computational model for Boltzmann entropy. This paper will review the evolutional processes, examine the bottleneck problems and the solutions, and finally propose a framework for the information theory of cartography. It is expected that such a theory will become the most fundamental theory of cartography in the big data era.

1. INTRODUCTION

Map is an effective communication means. It carries and transmits spatial information about spatial objects and phenomena, from map makers to map users. Therefore, cartography can be regarded as a communication system. According to Kent (2018), Keates (1964) "had introduced the idea of map as communication device at a London Meeting in 1964", but "the first published map communication model was devised by Moles (1964) and followed by Board (1967)". Since then, a number of cartographic communication models have been established. The typical example is the one developed by Kolacny (1969), as shown in Fig. 1.

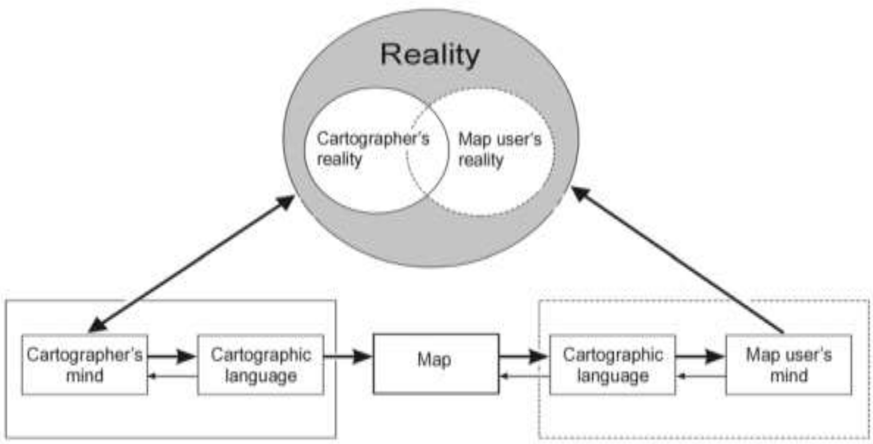

Fig. 1 The cartographic information transmission model proposed by Kolacny (1969)

Since 1960s, cartographers (e.g. Sukhov 1967, Knopfli 1983) made efforts on application of the information theory originated by Shannon (1948) and developed in digital communication community to cartography so as to develop a new theory of cartographic communication, or Cartographic Information Theory. However, it has been not very successful. It has been found that the bottleneck of the problem is the measures for map information. That is, Shannon entropy is only able to characterize the statistical information of map symbols but not capable of characterizing the spatial configuration (patterns) of map symbols.

It can be noted here that some break-through has been made and it is feasible to build such a theory now. The bottleneck problem is examined in Section 2, the break through is reported in Section 3, a framework is proposed in Section 4. Finally some remarks are made in Section 5.

2. LIMITATIONS OF SHANNON ENTROPY: ONLY FOR STATISTICAL INFORMATION

It has been found that the bottleneck of the problem is the measurement of map information. The Shannon entropy is the measure of information used in information theory. It is a statistical measure as follows:

$$H(X) = -\sum_{i}^{n} P(x_i) \times log\ P(x_i) \qquad (1)$$

Where, $P(x_i)$ is the probability of a random variable X taking a vale of x_i. In the case of maps, the X is map symbol, x_i is the i[th] type of symbol and n is the total number of symbol types. The proportion of the i[th] type of symbol to the total number of symbols on the map.

Fig. 2 shows two maps with quite different spatial configurations, but identical value of Shannon entropy is obtained from Equation (1), because both of them have the same types of symbols and have the same number of symbols in each type. It means that the Shannon entropy fails to characterize the configurational information of spatial objects on maps.

The Shannon entropy also fails to capture the configuration information of images. Fig. 3 shows three images with quite different spatial configurations, but identical values of Shannon entropy (H=3.2 bits) are obtained from Equation (1) because the middle and right images are the randomized result of the left image, thus with the same number of pixels for each gray value (ranging from 0 to 255).

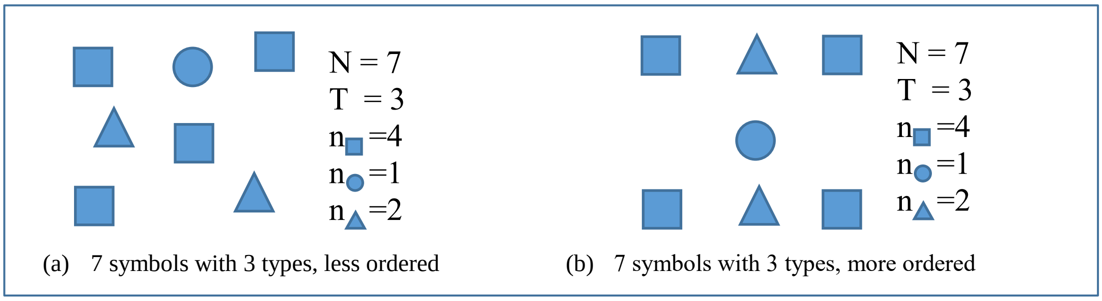

Fig. 2 Two maps with different distributions of symbols but same amount of Shannon entropy (i.e. H=0.415)

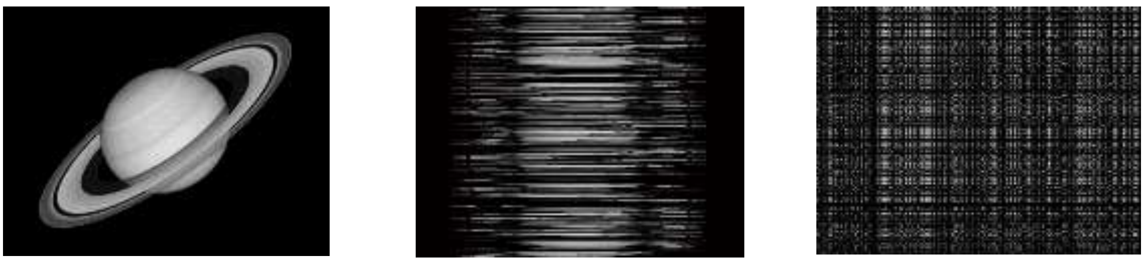

Fig. 3 Three images have very different spatial configurations but identical Shannon entropy (Gao and Li 2018a)

3. ENTROPY MODELS FOR CONFIGURATIONAL INFORMATION SUCCESSFULLY BUILT

However, the situation has been dramatically changed in recent years and the formation of Information Theory of cartography, or Cartographic Information Theory, is now possible. Because the solutions for the measurement of the configurational information of both graphic maps and image maps have been developed already.

Li and Huang (2002) have argued that, in addition to statistical information, the (geo)metric information, thematic information and topological information are all the essential properties of maps. As a result, they (Li and Huang 2002) developed mathematical models for these three types of map information. They adopted an idea of an influence region for each map symbol, which is defined as its Voronoi region. The Voronoi regions of all map symbols together form a Voronoi diagram, which is a tessellation (see Lee et al. 2000) of the map space. In this way, the disjoint map features are contiguously connected. Fig. 4 shows the space tessellation by the Voronoi diagram of map symbols. Naturally, the probability in Equation (1) is replaced by the proportion of the Voronoi region to the whole map space. Therefore, the metric information, donated as $H(M)$, is defined as follows:

$$H(M) = -\sum_{i=1}^{N}\left(\frac{S_i}{S}\right) \times log\left(\frac{S_i}{S}\right) \qquad (2)$$

Where, S_i $(i = 1, 2, \dots\dots, N)$ is the Voronoi region of the i^{th} map symbol; S is the whole map space (i.e. $S = \sum S_i$); and N is the total number of map symbols.

Let the i^{th} map symbol has a total of N_i symbols in its 1-order neighbourhood (with directly adjacent Voronoi regions, see Figure 4), belonging to M_i thematic types. The number of the j^{th} (j=1, 2, … M_i) type features is n_j, then the proportion for j^{th} type symbols is as follows:

$$P(x_j) = \frac{n_j}{N_i} \qquad (5)$$

Then, the entropy of thematic information for the i^{th} map symbol is as follows:

$$H_i(T) = -\sum_{j=1}^{M_i} P(x_j) log P(x_j) = -\sum_{j=1}^{M_i} \frac{n_j}{N_i} log \frac{n_j}{N_i} \qquad (6)$$

Therefore, the entropy for the thematic information of the whole map is as follows:

$$H(T) = \sum_{i=1}^{N} H_i(T) = -\sum_{i=1}^{N}\sum_{j=1}^{M_i}\left\{\frac{n_j}{N_i} log \frac{n_j}{N_i}\right\} \qquad (7)$$

In the example of Figure 4, the upper map has more thematic variety in the neighbourhood of each symbol, thus has a higher thematic information (i.e. H(T) =28.2) than the lower map (i.e. H(T) =16.4). It is also shown that the two maps have different metric information although they have identical statistical information.

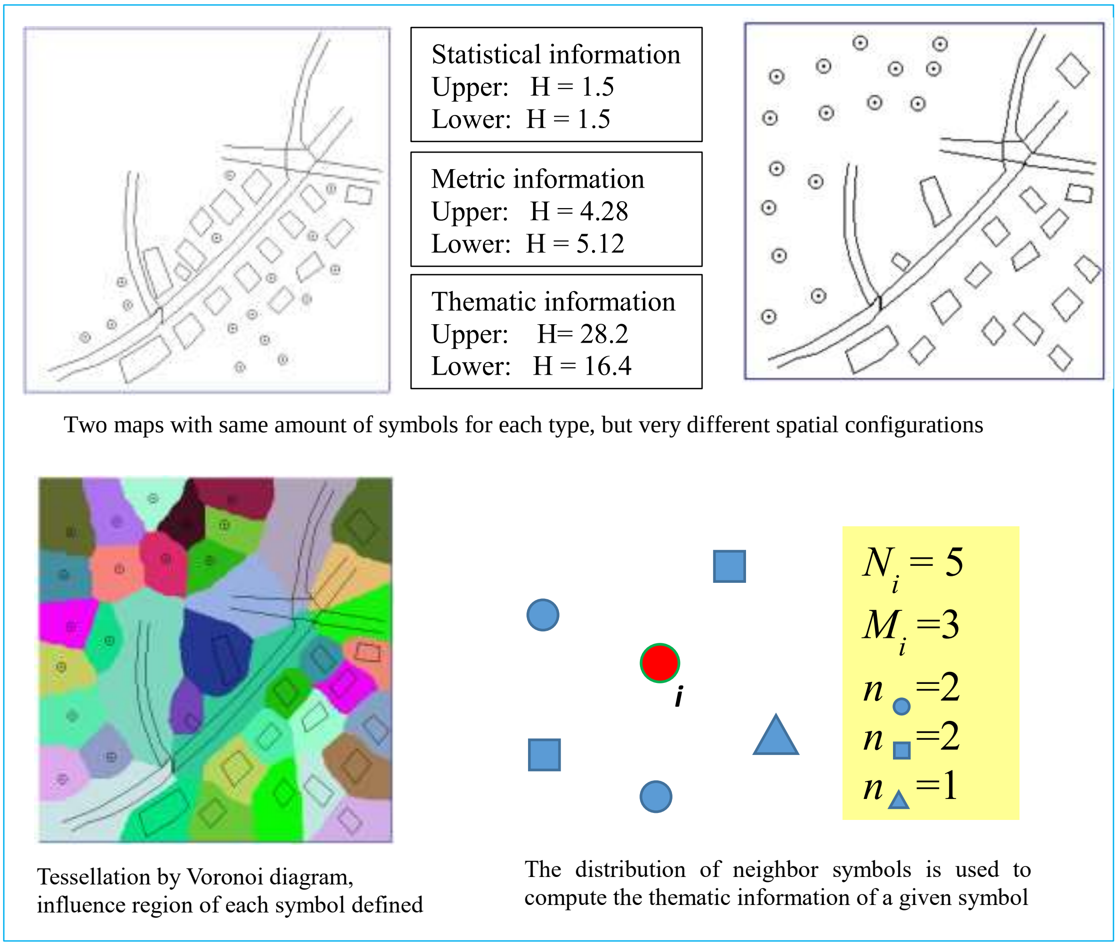

Fig. 4 Two maps with very different spatial configurations (modified from Knopfli 1983, and Li and Huang 2002), thus with very different configurational information, but identical Shannon entropy

To capture the configurational information of image (maps), a number of improved Shannon entropies and variants have been developed. However, a recent thermodynamics-based evaluation by Gao et al (2018), making use of validity, reliability and ability to capture the configurational disorder as metrics, reveals that these entropies are not thermodynamically sound. Therefore, calls have been made (e.g. Li 2017) to go back to Boltzmann entropy which is capable of characterizing the configurational entropy of images but has no computational solutions developed since 1872. The equation of Boltzmann entropy had been engraved on the Boltzmann tombstone in Vienna Central Cemetery (see Fig. 5), and is as follows:

$$S = k \log W \tag{8}$$

where, S is the Boltzmann entropy; k is the Boltzmann constant (as 1 in the case of digital images and landscape models as suggested by Cushman (2016) and W is the possible number of miscostates for a given macrostate.

Indeed, Gao et al (2017a) have developed the only feasible solution for the computation of Boltzmann entropy. They introduced multi-scale concepts into the system to define the two basic concepts, i.e. macrostate and microstate. For a given image, the macrostate is an image with a step of up-scaling, and the microstates are all the possibilities of down-scaling from the macrostate. As shown in the example in Fig.5, a 2x2 image is aggregated into 1x1 pixel to form a macrostate, then 4 microstates in one case and 6 microstates in other case can be obtained. Then computational algorithms gave also been developed for the computation of microstates (e.g. Gao et al 2017b, Gao and Li 2019a). It has been demonstrated by Gao and Li (2019b) that Boltzmann entropy can be used for the measurement of the spatial information of not only images but also maps.

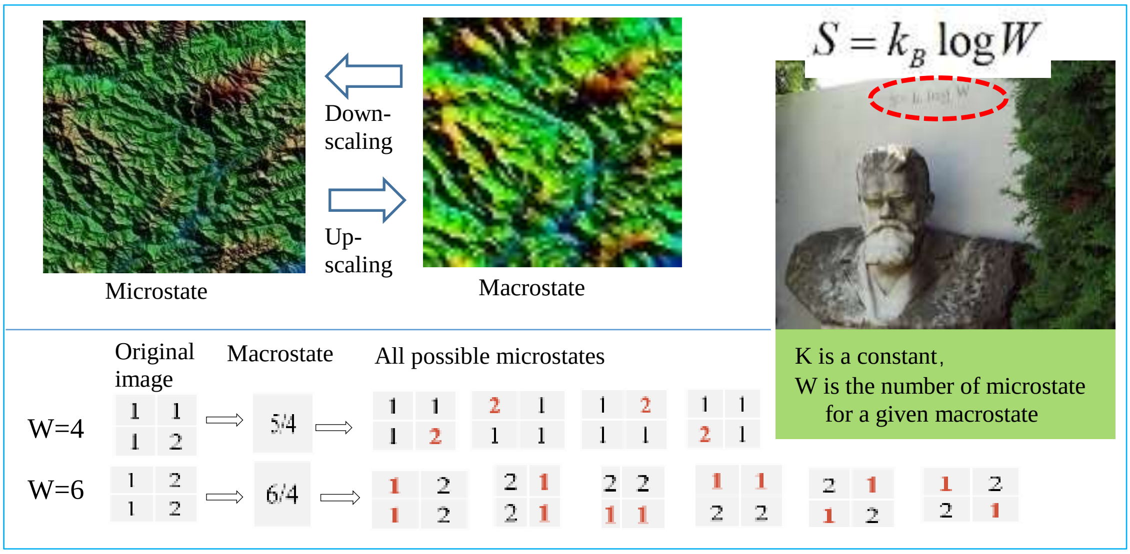

Fig.5 Definitions of macrostate and microstate for Boltzmann entropy (modified from Gao et al 2017)

4. A FRAMEWORK FOR INFORMATION THEORY OF CARTOGRAPHY

As information theory is defined as the mathematical theory concerned with the content, transmission, storage, and retrieval of information, usually in the form of messages or data, and especially by means of computers, the Information Theory of Cartography can be defined as the mathematical theory concerned with the content, transmission, storage, and retrieval of map information (or cartographic information), usually in the form of graphics and/or images, and especially by means of computers. With this definition, the objectives the theory are very clear.

The flow of information transmission in classic information theory is as shown in Fig. 6. By comparison with Fig. 1, it can be found that the recognition (of the reality and maps) component in the traditional cartographic communication (see Fig.1) cannot be included. Indeed, this component deals with information at the level of semantics but the Shannon deals information at syntax level. Therefore, the topics constituting the framework of The Information Theory of Cartography should include the following components:

- Measuring spatial information of graphic maps with generalized Shannon entropy;
- Measuring spatial information of image maps with Boltzmann entropy;
- Storage of map information;
- Information transmission from images to maps;
- Information transmission from one scale to another;
- Information change in transmission/display-mode; and
- Entropy-based optimization for map design.

A lot of research still needs to be carried out to make the Information Theory of Cartography become useful guide for cartographic practice although some work has already been done, such as transfer of information from a stage to another (e.g. Li and Huang 2001, Ai et al 2015) and optimal map design (e.g. Wang et al 2009, Bjørke 2012). Indeed, a systematic study on different topics as listed in this section is still a matter of urgency and this needs joint efforts from researchers in the geo-information communities such as ISPRS and ICA.

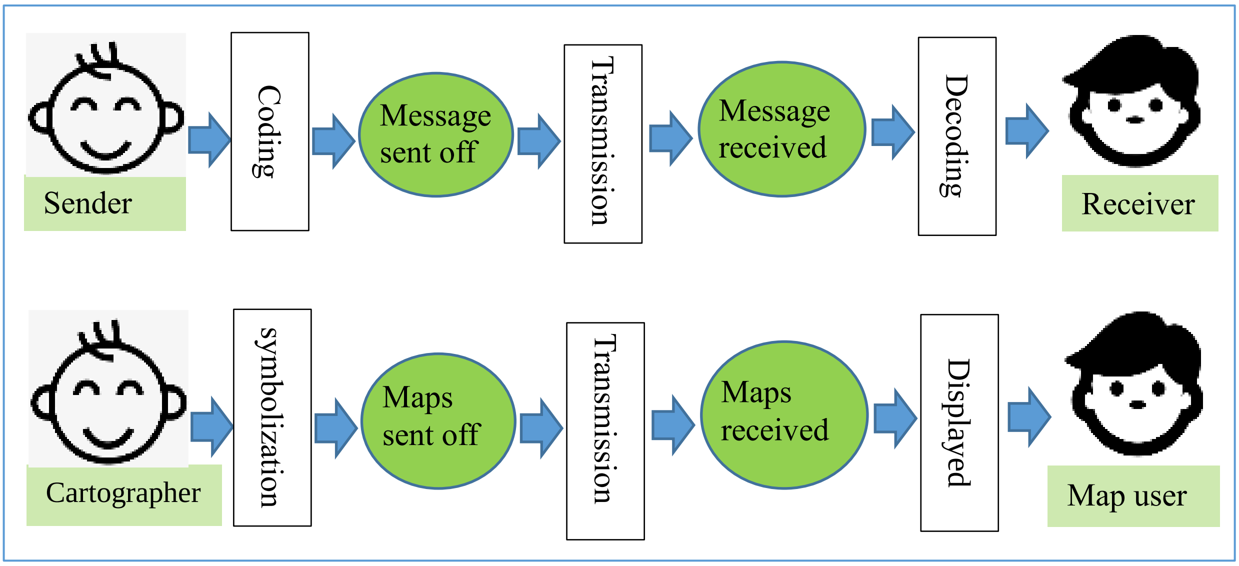

Fig. 6 The flow of information transmission in Cartographic Information Theory

5. CONCLUDING REMARKS

This paper deals with the construction of a framework for the Information Theory of Cartography (or Cartographic Information Theory) and proposal of a research agenda. It has been emphasized that

- there is a need of information-based theory for cartographic communication studies;
- the bottleneck problem in the building of such a theory is the appropriate measurement of the spatial (configurational) information of graphic and images maps;
- break-though has been made in the development of such measures; and
- it is now feasible to construct a framework for the Information Theory of Cartography and such a framework should include different aspects in the information flow from data to map and finally to map use.

It is expected that this paper will stimulate the discussions on this topic and a new branch of cartography, the Information Theory of Cartography, or Cartographic Information Theory, will become established.

It is also expected that such a theory will become the most fundamental theory of cartography in the big data era, because as advocated by Wheeler (1990) of Princeton University, the physical world is made of information, with energy and matter as incidental.

It is noted here that the curent study takes consideration of map information at syntax level only. Information at semantic and pragmatic levels should be studied in the future to make a Generalized Information Theory of Cartography established.

ACKNOWLEDGEMENTS

This work is supported by a NSFC key project (#41930104) and a Hong Kong RGC GRF project (#152219/18E).

REFERENCES

Ai T.H., He Y.K. and Du X., 2015. Information entropy change in GIS data scale transformation. *Geography and Geo-Information Science*, 31(2): 7-11.

Bjørke J.T., 2012. Exploration of information theoretic arguments for the limited amount of information in a map. *Cartography and Geographic Information Science,* 39(2): 88-97.

Board C., 1967. Maps as Models. In: Chorley R.J., and Haggett P. (eds.), *Models in Geography*. London: Methuen.

Cushman S.A. 2016. Calculating the configurational entropy of a landscape mosaic. *Landscape Ecology,* 31(3), 481-489.

Gao P.C. and Li Z.L., 2019a. Aggregation-Based Method for Computing Absolute Boltzmann Entropy of Landscape Gradient with Full Thermodynamic Consistency. *Landscape Ecology*, 34(8):1837-1847.

Gao P. and Li Z.L., 2019b. Computation of the Boltzmann entropy of a landscape: a review and a generalization. *Landscape Ecology,* 34 (9): 2183-2196.

Gao P. and Li Z.L., 2019b. Computation of the Boltzmann entropy of a landscape: a review and a generalization. *Landscape Ecology,* 34 (9): 2183-2196.

Gao P.C., Zhang H. and Li Z.L., 2017a. A hierarchy-based solution to calculate the configurational entropy of landscape gradients, *Landscape Ecology,* 32 (6), 1133-1146.

Gao P.C., Zhang H. and Li Z.L., 2017b. An Efficient Analytical Method for Computing the Boltzmann Entropy of a Landscape Gradient. *Transactions in GIS*, 22(5): 1046-1063.

Keates, J.S., 1964. Cartographic information. *Abstract of Paper Presented at the 20th International Geographical Congress.* London.

Kent A., 2018. Form follows feedback: Re-thinking cartographic communication. *Westminster Papers in Communication and Culture*, 13(2): 96-112.

Kolacny A., 1969. Cartographic information – a fundamental term in Modern Cartography. *The Cartographic Journal*, 6: 47-49.

Knopfli R. 1983. Communication Theory and Generalization. In: Taylor D R F. (ed.), *Graphic Communication and Design in Contemporary Cartography*. New York and Chichester: John Wiley & Sons, 1983: 177-218.

Li Z.L., 2017. Entropy for Image Quality Measure: From Shannon Back to Boltzmann. *Keynote Speech at the International Workshop on Image and Data Fusion*, *Geospatial Week 2017*, 18-22 September 2017, Wuhan, China.

Li Z.L. and Huang P.Z., 2001. Transformations of Spatial Information in Multi-scale Representation. *20th International Cartographic Conference*, 26-28 August 2001, Beijing. (CD-Rom)

Li Z.L. and Huang P.Z., 2002. Quantitative Measures for Spatial Information of Maps. *International Journal of Geographical Information Science,* 2002, 16(7): 699-709.

Moles A. A. 1964. Théorie de l'information et message cartographique. *Sciences et Enseignement des Sciences*, 5(32), 11–16.

Sukhov V I. Information Capacity of a Map Entropy. *Geodesy and Aerophotography*, 1967, 10(4): 212-215.

Wang Z., Wu Z.H., Fei L.F. and Li L.X., 2009. Automatic name placement of area feature: a metric information approach. *Acta Geodaetica et Cartographica Sinica*, 38(2): 183-188.

Wheeler J., 1990. Information, physics, quantum: the search for links. In: ZUREK, W. ed. *Complexity, Entropy, and the Physics of Information*. Boston, MA: Addison-Wesley.