Physica A 469 (2017) 572–587

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

An efficient link prediction index for complex military organization

Changjun Fan [a,*], Zhong Liu [a], Xin Lu [a,b], Baoxin Xiu [a], Qing Chen [a]

[a] Science and Technology on Information Systems Engineering Laboratory, National University of Defense Technology, Changsha Hunan 410073, China

[b] Department of Public Health Science, Karolinska Institute, Stockholm 17177, Sweden

HIGHLIGHTS

- For the first time, we quantify the nodes' types effect on their linking behaviors, and empirically proved that it could remarkably improve the prediction accuracy of 25 current methods.
- We design a new link prediction index for heterogeneous military network and it is superior to all the other methods both in missing links prediction and spurious links identification tasks.
- We investigate the algorithms' robustness under noisy environment, and demonstrate that our method maintains the best performance under the condition of small noise.

ARTICLE INFO

Article history:
Received 23 June 2016
Received in revised form 27 August 2016
Available online 18 November 2016

Keywords:
Complex military organization
Link prediction
FINC-E model
Social organization

ABSTRACT

Quality of information is crucial for decision-makers to judge the battlefield situations and design the best operation plans, however, real intelligence data are often incomplete and noisy, where missing links prediction methods and spurious links identification algorithms can be applied, if modeling the complex military organization as the complex network where nodes represent functional units and edges denote communication links. Traditional link prediction methods usually work well on homogeneous networks, but few for the heterogeneous ones. And the military network is a typical heterogeneous network, where there are different types of nodes and edges. In this paper, we proposed a combined link prediction index considering both the nodes' types effects and nodes' structural similarities, and demonstrated that it is remarkably superior to all the 25 existing similarity-based methods both in predicting missing links and identifying spurious links in a real military network data; we also investigated the algorithms' robustness under noisy environment, and found the mistaken information is more misleading than incomplete information in military areas, which is different from that in recommendation systems, and our method maintained the best performance under the condition of small noise. Since the real military network intelligence must be carefully checked at first due to its significance, and link prediction methods are just adopted to purify the network with the left latent noise, the method proposed here is applicable in real situations. In the end, as the FINC-E model, here used to describe the complex military organizations, is also suitable to many other social organizations, such as criminal networks, business organizations, etc., thus our method has its prospects in these areas for many tasks, like detecting the underground relationships between terrorists, predicting the potential business markets for decision-makers, and so on.

* Corresponding author.
E-mail address: fanchangjun09@163.com (C. Fan).

http://dx.doi.org/10.1016/j.physa.2016.11.097
0378-4371/© 2016 Elsevier B.V. All rights reserved.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

With the rapid development of information technology and military science, the state-of-art complex military organization named "System of System, SOS" has been playing an increasingly significant role in nowadays' warfare, disaster response, nation building, peace operations and counter-terrorism [1]. The complex military organization is often formed as a complex network, where nodes represents functional entities that are themselves complex, edges denotes various kinds of communication relationships, functional entities interact with each other to obtain the common shared goals [2].

Modeling complex military organization is a challenging task, due to the following three reasons. One is the complexity of the functional entities, these numerous entities are distributed in the multi-dimensional physical space, such as land, ocean, inner space, outer space, electromagnetic space and cyberspace, and in the virtual space like information space, cognitive space and social space; second is the complexity of dense, real time, and frequent relationships between entities, making the whole system nearly seamless; the last is that nearly all entities are involved with human, to model the human-level adaptability is very hard. Despite the above problems, there are still some inspiring works which motivate a more accurate model. Alberts et al. [3] pointed out the integrative system should produce and utilize the information superiority, and integrate the Command and Control units, weapon system and forces effectively to improve the ability of information sensing, intelligence sharing and coordination. In 2011, he proposed the concept of "Edge Organization" [4], and proved its agility advantages on the basis of complex military organization. Inspired from the recent advances of network science, network is confirmed to be an useful tool, which reflects the import characteristics of military organization model. These network models for military organization are not simply recognized as a communication network, but complex networks including command and control, society, environment, weapon, radar, etc [5]. Cares [6], Carley [7] and Krackhardt [8] all attempt to build network models to complex military organizations, however, their proposed nodes and edges are all homogeneous, which is not reliable in reality. Dekker [9–12] put forward the FINC model, and depicted the heterogeneous nodes, which shed new lights on the complex military organization. And Guoli Yang [13] added the heterogeneous edges in FINC model, and proposed an extended FINC model, named FINC-E, with weighted functional nodes and edges.

FINC (Force, Intelligence, Networking and C2) methodology classified the nodes into three types: C2 node (C2), like command post, control center, etc., Intelligence node (I), like radar, AWACS, etc., and Force node (F), like missile position; and links provide communications between nodes, indicated by lines or arrows, depending on whether information flow is bidirectional or unidirectional. Fig. 1 is an illustration for FINC model, where circle nodes represent C2 nodes, square nodes denote force nodes, triangle nodes indicate intelligence nodes. In FINC-E model, there are five types of links, Intelligence link (I $\rightarrow$ C2, unidirectional), C2 link (C2 $-$ C2, bidirectional), Fire link (I$-$ $>$ F, unidirectional), Decision link (C2-F, bidirectional) and Communication link (I–I, bidirectional). Each node has many attributes, like attack cost, InEdge, OutEdge, etc., each link also has attributes, such as information transfer delay, information load, information accuracy, attack cost, InNode, OutNode, etc. More details see in Ref. [13]. In this paper, FINC-E model is utilized to model the complex military organization.

As for the problem of link prediction, it was originated from computer science study, also known as link mining, and has been studied for a long time. Recent researches about link prediction on complex network [14] have drawn much attention, since it utilizes structural information only, and obtains satisfactory performance. Link prediction on complex network attempts to estimate the likelihood of the existence of links between nodes based on the attributes of nodes as well as the structure of networks [15,14,16]. It is divided into three categories, one is predicting missing links, i.e., existent yet unknown links; one is predicting future links, i.e., may exist or appear in the future of evolving networks; last is identifying spurious links, i.e., nonexistent yet observed links, also known as noise [17]. Due to its formal simplicity, theoretical value and practical significance, link prediction has attracted increasing attentions from various fields of researches and engineers, such as physicists, mathematicians, computer scientists, statisticians, biologists, etc.

As for complex military organization, link prediction may be more significant, since the war determines a country of vital significance, and quality of intelligence is critical to military decisions, as a formal representation of military intelligence, network topology for complex military organization has to be true enough to guarantee the reliability of the subsequent analysis, such as critical operational units analysis, community analysis and network evolving analysis. However, due to the complex battle field situations and the expensive costs of military intelligence collection, it is nearly impossible to obtain an absolute accurate military network, there must be some missing information or spurious noise, in other words, missing links and spurious links in the network topology. If we can predict them or identify them in advance with link prediction methods, it would be both meaningful to optimize the military organization structure for our side and attack the other side's critical operational components, which are sure to enhance the accuracy of military decisions and accelerate the process of victory.

Current link prediction methods are mainly designed basing on the definition of node similarity, which assumes that the greater the similarity values between nodes are, the higher the likelihood of the existence of links between them [18]. There are many methods to measure node similarity, and one of the simplest is calculated just by the observed node attributes, i.e., two nodes are defined to be similar if they share many common characters [19], Popescul and Ungar [20] have done

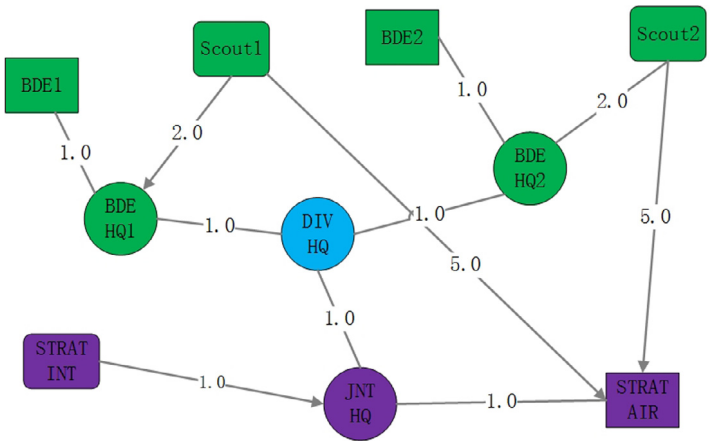

Fig. 1. FINC-E model for Air-Strike Scenario from Fig. 7 of Deckker (2001). This military network contains all kinds of nodes and edges, and their structure and capability are dependent on their roles and relationships in the organization.

some works on this aspect with machine learning or Markov process ways, however, these methods are limited due to the difficulty of obtaining the nodes attributes information. Another branch is based on the network structure only, which is named structural similarity, and can be further divided into three types: local information based similarity, such as CN index; path-based similarity, such as Katz index; random walk-based similarity, such as ACT index. There are 24 main similarity indexes now, details about them see in SI. And they are utilized to compare the prediction accuracy with the method proposed in this paper. The above methods are designed for homogeneous networks, and recent some studies have worked on the link prediction in heterogeneous networks [21–24], where there are different types of nodes and types. And there are two typical handling ways: one is treating all types of link equally; the other is studying each type of link independently and ignoring its correlation with other link types. However, both the above methods could not be applied directly in this paper. The first methods cannot handle the complex military network's heterogeneity, the second methods may lead to much information loss.

Complex military network is a special heterogeneous network, there are three different types of nodes, and linking behaviors are different between them, for example, if there are command nodes and intelligence nodes densely connected in a control-centered military organization, intelligence nodes are more likely to build links with commands nodes than forces nodes. As a result, an index combining this impact and traditional topological similarity is proposed in this paper to improve the link prediction accuracy in complex military organization network.

The major contributions of this study are summarized as follows:

1. For the first time, we quantify the nodes' types effect on their linking behaviors, and empirically proved that it could remarkably improve the prediction accuracy of all the current methods;
2. We design a new link prediction index for heterogeneous military network and it is superior to all the other methods both in missing links prediction and spurious links identification tasks;
3. We investigate the algorithms' robustness under noisy environment, and demonstrate that our method maintains the best performance under the condition of small noise, and we also observe that spurious links are more destructive than missing links in military networks, which is just opposite to that in recommendation systems.

Of separate interest is the reproducibility of our work. Dataset, source code and extra source code of competing methods can be found on our websites.[1]

The remainder of this paper is organized as: Section 2 describes the methodology in detail, Section 3 conducts experiments on a real military network data and compared the performance of our method with all the other methods, Section 4 concludes the paper and pointed the further directions.

2. Methodology

Complex military organization can be described as $G(V, E, A, L, W)$ in FINC-E model, where V denotes operation units, E indicates communication links between nodes, A is a set of node types, including three types here, command nodes, fire nodes and intelligence nodes, L represents link types, mainly refers to five types, Intelligence link (I $\rightarrow$ C2, unidirectional), C2 link (C2 $-$ C2, bidirectional), Fire link (I $\rightarrow$ F, unidirectional), Decision link (C2 $\rightarrow$ F, bidirectional) and Communication link (I $-$ I, bidirectional). W is a set of weights, including node weights, like attack costs, and link weight, such as information

[1] https://github.com/FFrankyy/Link-prediction-heterogeneous-social-organization.

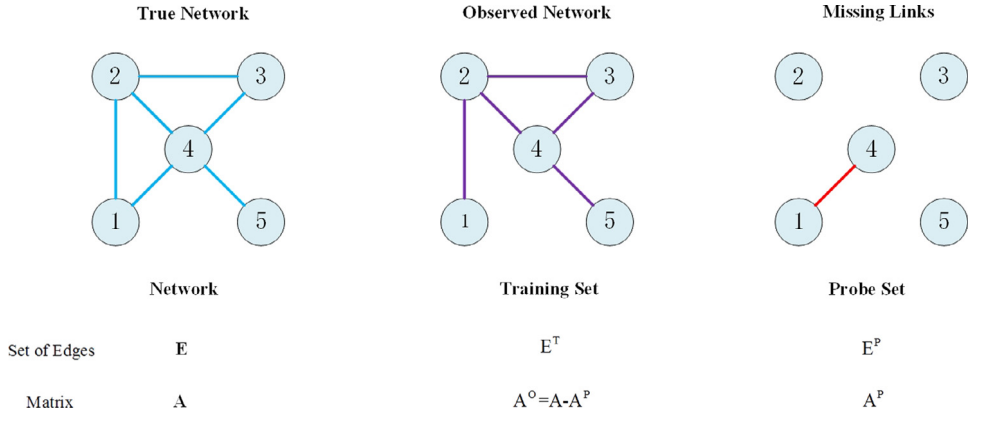

Fig. 2. Illustrating network $G(V, E)$ with $|V| = 5$ nodes and $|E| = 6$ links for predicting missing links.

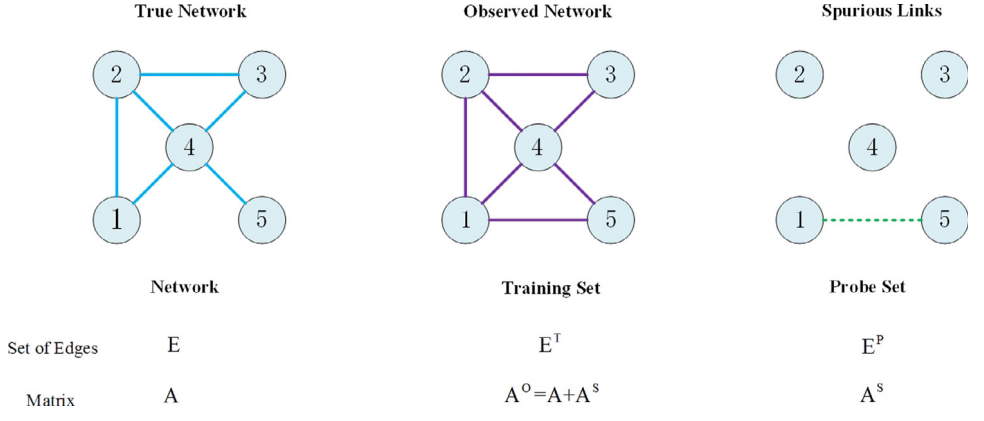

Fig. 3. Illustrating network $G(V, E)$ with $|V| = 5$ nodes and $|E| = 6$ links for identifying spurious links.

loads. Just as mentioned in Introduction section, due to the complex battlefield situations and expensive costs of intelligence collection, there are often filled with noisy or incomplete information in the observations, which are reflected as spurious links or missing links respectively in the military network.

To purify the network topology, we used link prediction methods to predicting the missing links and identifying the spurious links based on the known network topology and nodes types information.

For the missing link prediction, the task is to estimate the existence tendency of all the non-observed links. Specifically, consider an network $G(V, E)$, where V is the set of $|V|$ nodes and E is the set of $|E|$ links, U is all possible links, the number is $|V|(|V| - 1)/2$, then the nonexistent links is $U - E$. Assume there are some missing links (or the future links) in the set $U - E$, and the task of link prediction is to find out these links. Generally we have no ideas about which links are the missing or future ones, otherwise we do not need to do predictions. To test the algorithms' accuracy, the observed links, E, is randomly divided into two parts: the training set, E^T is treated as known information, while the probe set, E^P is used for testing. $E^T \bigcup E^P = E$, $E^T \bigcap E^P = \emptyset$. In Fig. 2, the true network contains five nodes and six links, while the link (1,4) is missing in the observed network A^O. Then the missing link constitutes the probe set E^P, and the observed links constitute the training set E^T.

For spurious link identification, the task is to evaluate the reliability of all the observed links. Specifically, consider an network $G(V, E)$, where V is the set of nodes and E is the set of links. The set of observed links is E. Assume that there are some spurious links in the set E, the task of spurious link identification is to find out these links. To test the algorithm's accuracy, we randomly add some nonexistent links which constitutes the probe set E^P, and the given network (true network) together with the probe set constitute the training set E^T. Clearly, $E^T - E^P = E$, $E^T \bigcap E^P = E^P$. In Fig. 3 the true network contains five nodes and six links, the spurious link (1,5) was added to the network to construct the training set E^T, while the spurious link constitutes the probe set E^P.

Traditional methods for predicting missing links and identifying spurious links can be roughly divided into two classes: the probabilistic models and the similarity based algorithms: the former usually requires much information about node attributes, in addition to the observed network topology; the latter assigns a similarity score to every pair of nodes and ranks all links according to their scores. Since it is always very hard to collect enough nodes attributes, similarity based algorithms

are thus the practical option, how to define the similarity between nodes is the key point for similarity algorithms, here we proposed the *LR* (Link Reliability) index, combining both the nodes' types information and their structural similarities.

2.1. Design of LR (Link Reliability) index

Consider a complex military organization network $G(V, E, A, L, W)$, it is a heterogeneous directed weighted network, and it is difficult to directly conduct link prediction tasks, here we simplified it just a heterogeneous undirected unweighted network based on the following considerations: (1) link directions are totally determined by the two nodes linked, if two nodes' types are known, links between them are determined, thus links could be regarded as undirected; (2) weights in the complex military organization network are hard to obtain, and they have no obvious effects on linking behaviors between nodes, thus the weights could be regarded as the same 1 in the network. As a result, in this paper, link prediction is actually conducted in the network $G'(V, E, A)$, and our index, named Link Reliability (*LR*) is designed for it.

As we analyzed before, whether two nodes in the military network have links or not depends on two parts, one is the two linked nodes' types, denoted as T, the other is these two nodes' topological similarity, denoted as S. Let R stands for the link reliability, then we may have the following:

$$R = f(T, S) \tag{1}$$

f is the combined function of these two parts' effects.

Next, we deduce the forms of effect of nodes' types T, effect of node structural similarity S and the combined function f.

2.1.1. Effect of nodes' types

There are three types of nodes in the military network, and linking behaviors between different types of nodes are different, how to quantify this effect? Here, the stochastic block model is utilized to give the mathematical formula of it.

Stochastic block model is a general network model [25], it divides the nodes in the network into several groups, and the existence tendency between nodes is determined by the groups they belong to, nodes in the same group have the same linking behaviors. The model is most suitable for the situations where nodes' types significantly affect their linking behaviors, which is just the case this paper is to address.

A stochastic block model is composed of two parts: one is all the possible group division strategies, denoted as Ω; the other is the linking probability matrix between different groups, denoted as Q. Given a specific block division P and the corresponding linking probability matrix Q, a stochastic block model $M = (P, Q)$ is then determined. The observed network A^O could be regarded as one from an unknown stochastic block model, let Ψ be a certain network property, then it is easy to obtain the following based on Bayes theorem:

$$p(\psi|A^O) = \int_{\Theta} p(\psi|M)p(M|A^o)\mathrm{d}M \tag{2}$$

where Θ stands for all possible stochastic block models, and since $p(M|A^O)p(A^O) = p(A^O|M)p(M)$, formula (2) could be rewritten as:

$$p(\psi|A^o) = \frac{\int_{\Theta} p(\psi|M)p(A^o|M)p(M)\mathrm{d}M}{\int_{\Theta} p(A^o|M')p(M')\mathrm{d}M'}. \tag{3}$$

Let Ψ be $A_{xy} = 1$, then $p(\Psi|A^O) = p(A_{xy} = 1|A^O)$ represents the linking reliability of node pair v_x, v_y based on the observed network A^O. Now we can obtain the form of T by proving the following Theorem 1. Detailed proof sees the Appendix A.

Theorem 1. *Given a complex military organization network $G(V, E, A)$, let σ_x denotes the type of node v_x, σ_y denotes the type of node v_y, and r_{σ_x,σ_y} is the number of all possible links between these two types of nodes. $l^O_{\sigma_x,\sigma_y}$ represents the observed links between these two types nodes. Then the linking reliability for node pairs $\{v_x, v_y\}$ could be calculated as:*

$$p(A_{xy} = 1|A^O) = \frac{l^O_{\sigma_x,\sigma_y} + 1}{r_{\sigma_x,\sigma_y} + 2}. \tag{4}$$

Therefore, formula of T is represented as $T(v_x, v_y) = p(A_{xy} = 1|A^O) = \frac{l^O_{\sigma_x,\sigma_y}+1}{r_{\sigma_x,\sigma_y}+2}$.

2.1.2. Effect of nodes' structural similarity

Military network is a sparse network, where there are abundant nodes while with a few links, this may be the result of the hidden hierarchical structure in military organization. In Ref. [14], some random walk based indices, such as *RWR*, *SRW*,

have been empirically demonstrated to be superior to the other methods on space networks, like the Power network and the Router network. As a result, we utilize the *RWR* indices, which obtained the best performance among all random walk based methods [14], to measure the effect of nodes' structural similarity in the military network.

RWR index is a direct application of the PageRank algorithm [26]. Consider a random walker starting from node v_x, who will iteratively move to a random neighbor with probability c and return to v_x with probability $1-c$. Denote $\overrightarrow{q}_x$ as the steady state of this random walker from node v_x, we have:

$$\overrightarrow{q}_x = cP^T\overrightarrow{q}_x + (1-c)\overrightarrow{e}_x \tag{5}$$

where P is the transition matrix with $P_{xy} = 1/k_x$ if v_x and v_y are connected, and $P_{xy} = 0$ otherwise, $\overrightarrow{e}_x$ is the vector with xth element equals 1, and the left elements are all 0s. It can be obviously obtained as follows:

$$\overrightarrow{q}_x = (1-c)(1-cP^T)^{-1}\overrightarrow{e}_x. \tag{6}$$

The *RWR* index is thus defined as

$$s_{xy}^{RWR} = q_{xy} + q_{yx} \tag{7}$$

where q_{xy} is the yth element of the vector $\overrightarrow{q_x}$. A fast algorithm to calculate this index was proposed by Tong et al. [27], and the application of this index to recommender systems can be found in Ref. [28].

Thus the formula of S is $S(v_x, v_y) = s_{xy}^{RWR} = q_{xy} + q_{yx}$.

2.1.3. Form the combined function f

Here we just utilize the simplest weighted average form to combine the two parts, which means the Link Reliability value could be calculated as:

$$R = f(T, S, \lambda) = \lambda T + (1-\lambda)S = \lambda T(v_x, v_y) + (1-\lambda)S(v_x, v_y). \tag{8}$$

And $R(v_x, v_y) = \lambda T(v_x, v_y) + (1-\lambda)S(v_x, v_y)$, form of $T(v_x, v_y)$ and $S(v_x, v_y)$ see formula (4) and (7). Weight parameter λ is proportional to each part's prediction accuracy in specific cases. For instance, for a specific network A, if the first part's accuracy is T_A^{AUC}, the second part's accuracy is S_A^{AUC}, then we obtain:

$$\frac{\lambda}{1-\lambda} \approx \frac{1-f(T,S,0)}{1-f(T,S,1)} = \frac{1-S_A^{AUC}}{1-T_A^{AUC}} \Rightarrow \lambda \approx \frac{1-S_A^{AUC}}{2-(T_A^{AUC}+S_A^{AUC})}. \tag{9}$$

2.2. Evaluation metrics

Two evaluation metrics adopted here are *AUC*(Area Under the receiver operating characteristic Curve) [29] and *R* [30]: *AUC* is used to measure the algorithms' accuracy in predicting missing links and identifying spurious links, and *R* is utilized to measure the ability for algorithms to maintain the prediction accuracies under noisy environments.

2.2.1. AUC

The *AUC* is a way to quantify the accuracy of prediction algorithms. Denote U as all the possible links in the network, E^T is the training set, E^P is the probe set. Given the rank of all non-observed links based on the descending similarity scores, the *AUC* value can be interpreted as the probability that a randomly chosen missing link (i.e., a link in E^P) is given a higher score than a randomly chosen nonexistent link (i.e., a link in $U - E^T$). At each time, we randomly pick a missing link and a nonexistent link to compare their scores, if among n independent comparisons, there are n' times the missing link having a higher score and n'' times they have the same score, the *AUC* value for the missing links prediction task could be calculated as:

$$AUC = \frac{n' + 0.5n''}{n}. \tag{10}$$

The above metric can also be used to quantify the performance on identifying spurious links. In such a case, a number of spurious links are randomly generated to constitute the probe set E^P, the observed links together with E^P constitutes the training set E^T. In contract to the predicting algorithm, the identification algorithm sort the links in E^T in the ascending order according to their similarity scores. The *AUC* value in this task becomes the probability that a randomly chosen links in E^P (i.e., a spurious link) is ranked lower that a randomly chosen spurious link (i.e., a link in E^P) is ranked lower than a randomly chosen existent link (i.e., a link in $E^T - E^P$). At each time, we randomly pick a spurious link and a existent link to compare their scores, if among n independent comparisons, there are n' times the spurious link having a lower score and n'' times they have the same score, the *AUC* value for the spurious links identification task could also be calculated as formula (10).

If all the scores are randomly generated from an independent and identical distribution, the *AUC* values should be about 0.5. Therefore, the degree to which the *AUC* value exceeds 0.5 indicates how better the algorithm performs than pure chance.

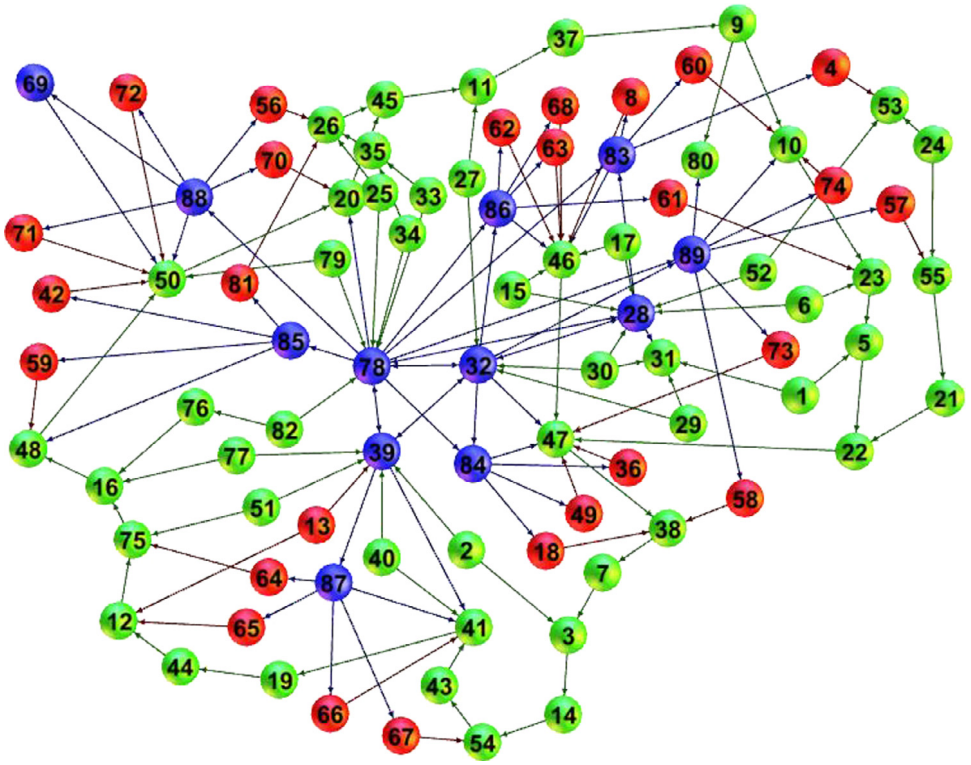

Fig. 4. Topology of the military organization network. Blue nodes represents the C2 units, like command posts, control center, etc., red nodes denotes the Force units, such as the missile position, and green nodes indicates the Intelligence units, like various optical stations, radar stations. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2.2.2. R

The *R* metric is recently proposed by Peng Zhang et al. [30] to quantify to what extent the link prediction algorithms can resist the noise in the observed network. Since the traditional link prediction algorithms are often conducted on the assumed noisy-free data (here refer to the training set), but what if the training set is incomplete or have some noise? How would the prediction accuracy be affected? R is such a metric to evaluate algorithm's robustness under noisy conditions. Mathematically, it reads:

$$R = \frac{1}{|L|}\sum_{q=0}^{|L|}\frac{AUC(q)}{AUC(0)} \tag{11}$$

where $L = ratio * |E^T|$, and *ratio* is a quantity to measure the fraction of randomly added or deleted links. When *ratio* is positive, $|ratio| * E^T$ links are randomly added to the training set, when *ratio* is negative, $|ratio| * E^T$ links are randomly deleted from the training set. In order to keep the network connected, we cannot remove too many links, as a result, we keep $-0.3 \leq ratio \leq 1$, $AUC(q)$ is the *AUC* value of a link prediction method when q links are added ($ratio > 0$) to or deleted ($ratio < 0$) from the training set, when $ratio = 0$, $R = 1$.

3. Experiment design and result analysis

3.1. Statistical descriptions for the experimental dataset

Since military data is of high secret, there are few public related data, as a result, there lacking enough validations for our method, which may make the results obtained in this paper bit of unconvincing, thus more experiences should be further explored. In this paper, we provide a real military network, which is from an area's military organization (specific name is anonymized here due to the requirement of secret protections, but detailed network information is described in S1), it contains 89 entities, including 12 command nodes, 26 force nodes, and 51 intelligence nodes, and there are 150 observable links, including 30 Intelligence links (I $\rightarrow$ C2, unidirectional), 16 Fire links (I $\rightarrow$ F, unidirectional), 26 Decision links (C2 $-$ F, bidirectional), 51 Communication links (I $-$ I, bidirectional) and 17 C2 links (C2 $-$ C2, bidirectional). The topology of the network is drawn as Fig. 4.

Table 1
Basic topological features of the military network. $|V|$ and $|E|$ are the number of nodes and links. F,I,C denotes the node number of the corresponding types, ρ is the network density, C2 is the clustering coefficient and r is the assortative coefficient. $\langle k\rangle$ is the average degree, $\langle d\rangle$ is the average shortest distance, and H is the degree heterogeneity, as $H = \langle k^2\rangle/\langle k\rangle^2$.

$\|V\|$	$\|E\|$	F	I	C2	ρ	C	r	$\langle k\rangle$	$\langle d\rangle$	H
89	150	26	51	12	0.038	0.094	0.2754	3.483	8.113	1.6337

Table 2
Top 10 nodes for five centrality values.

Index	1	2	3	4	5	6	7	8	9	10
Degree	78	46	47	50	89	32	86	88	41	83
Closeness	78	28	39	89	32	88	86	87	83	84
Betweenness	24	57	35	56	60	62	63	68	74	88
PageRank	47	28	39	46	50	32	10	41	26	20
Eigenvector	17	15	30	6	29	25	1	24	34	33

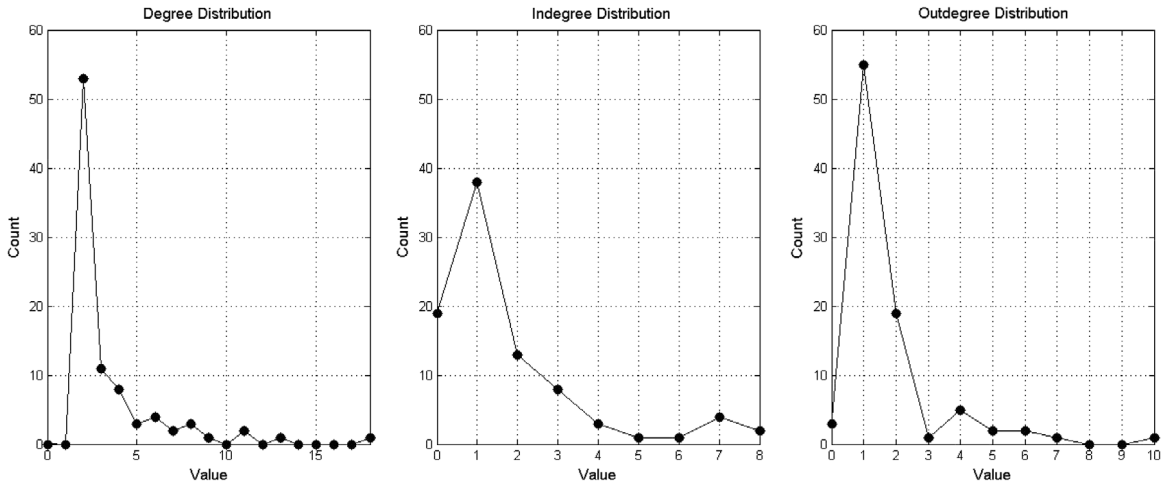

Fig. 5. Degree distribution of the military network. Value is the degrees, Count is the number of the corresponding degree nodes.

Now we demonstrate the statistical descriptions for the military network, from the perspective of basic topological features, degree distributions and centrality analysis.

3.1.1. Basic topological features

The basic topological features are listed as Table 1.

As can be seen from Table 1, the military network is a very sparse, loosely connected and assortative network, the average distance between nodes is 8.113, and the variance of nodes' degrees is 1.6337.

3.1.2. Degree distributions

Analyze the degrees for each node, we find that node 78 has the biggest degree 18, and it is the most important control center in the real military organization; the smallest degree is 2, and the average degree is 3.5. The degree distribution (including in-degree distribution and out-degree distribution) is plotted as Fig. 5.

As can be observed from Fig. 5, the total degree distribution, in-degree distribution and the out-degree distribution all obey an approximate power-law distribution, indicating that the majority of nodes have small degrees, while there are still a few nodes with very huge degree, like node 78 here.

3.1.3. Centrality analysis

Next, we consider the node centralities. There are many centrality indexes in complex networks, each indices measure the node importance from a specific perspective. Three most well-known centrality indexes are degree centrality, closeness centrality and betweenness centrality, here we add two more indexes, PageRank centrality and eigenvector centrality. List top 10 nodes for each centrality values in Table 2.

It can be seen from Table 2 that node 78, the most important control center, has the biggest degree, means it connects with the most neighbors, it is also the one closest to the other nodes (closeness centrality); node 24 has the biggest betweenness

Table 3
Comparisons for the missing link prediction accuracy measured by *AUC*. Each data point is obtained by averaging over 1000 independent implementations, and the bold number emphasizes the highest value. The best parameters for some indexes are: *LR*(0.3), *Katz*(0.01), *LHNII*(0.9), *RWR*(0.95), *LRW*(5), *SRW*(5), *SimRank*(0.8). The variance value should multiply e−2 actually.

Index	*LR*	*CN*	*Salton*	*Jacca*	*Soren*	*HPI*	*HDI*	*LHN*	*AA*
AUC	**0.7741**	0.6081	0.6065	0.6044	0.6044	0.6092	0.6031	0.605	0.6087
Variance	0.3041	0.275	0.2724	0.2631	0.2631	0.285	0.2579	0.2692	0.2733
Index	*RA*	*PA*	*LNBCN*	*LNBAA*	*LNBRA*	*LocalP*	*Katz*	*LHNII*	*ACT*
AUC	0.6087	0.5043	0.6068	0.6068	0.6069	0.6303	0.618	0.5598	0.49
Variance	0.2733	0.4852	0.2777	0.2774	0.2778	0.4167	0.5806	0.4911	0.5087
Index	*Cos+*	*RWR*	*LRW*	*SRW*	*MFI*	*TSCN*	*TSAA*	*SimRank*	
AUC	0.7615	0.746	0.6958	0.6925	0.689	0.6007	0.6284	0.7341	
Variance	0.2436	0.3869	0.4559	0.4745	0.3972	0.593	0.5844	0.3365	

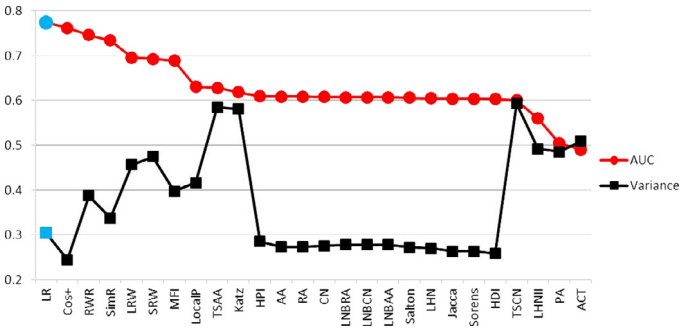

Fig. 6. Comparisons for the missing link prediction accuracy measured by *AUC*. Each data point is obtained by averaging over 1000 independent implementations, and the bold number emphasizes the highest value. The best parameters for some indexes are: *LR*(0.3), *Katz*(0.01), *LHNII*(0.9), *RWR*(0.95), *LRW*(5), *SRW*(5), *SimRank*(0.8).

value, indicating that it is located at the transportation center, controlling most of the shortest paths between any two nodes, actually it is an optical station, and many communications must transfer across it; node 47 has the biggest PageRank values, meaning that it is the most easily accessible node when starting from one node or could be regarded as the node most likely to be digged out, and it is also an optical station; node 17 has the highest eigenvector value, which means it has the most high quality neighbors, and it is a radar station in reality, and is linked with two control centers and one important optical station.

3.2. Missing links prediction

Divide the network into two parts, the training set E^T contains 90% of the links, and the remaining 10% of links constitute the probe set E^P. It is obvious that $E^T + E^P = E$, and $E^T \bigcap E^P = \emptyset$. Compared our proposed method *LR* with all the other similarity-based indexes (details see S2) on the performance of *AUC* (formula (10)) values. The prediction accuracies measured by *AUC* are shown in Table 3 and Fig. 6. Each data point is obtained by averaging over 1000 implementations with independently random divisions of the training set and the probe set, and the bold number in the table emphasizes the highest accuracy. It is obvious from the comparisons that our method performs the best among all state-of-the-art algorithms (about 1.26% higher than the current best method), and its variance value is medium among all, indicating it is also a comparatively stable index. In Fig. 7, we further demonstrate that such results are not sensitive to the size of the probe set.

3.3. Spurious links identification

Next we consider the identification of spurious links, where spurious links are those links being observed but not really existent, which may be resulted from experimental errors or data noise. To test the validity of the algorithms, we randomly add 10% ($0.1|E|$) nonexistent links to the network data which constitute the probe set E^P, and the original network together with spurious links constitute the training set E^T. It is obvious that $E^T \bigcap E^P = E^P$, $E^T - E^P = E$. Compared *LR* with other methods on the performance of *AUC* values, results are shown in Table 4 and Fig. 8. Each data point is averaged over 1000 independent runs with different randomly generated spurious sets. Again, our method is remarkably better than all other state-of-the-art methods, specifically, 3.05% higher than the current best method, and not sensitive to the size of probe set from Fig. 9.

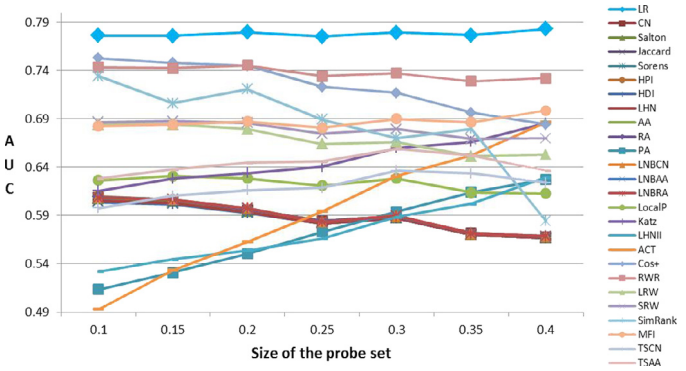

Fig. 7. Predicting missing links for different sizes of probe set. Size of the probe set ranges from 0.1 to 0.4. Also each data point in the figure are obtained by averaging over 1000 independent implementations.

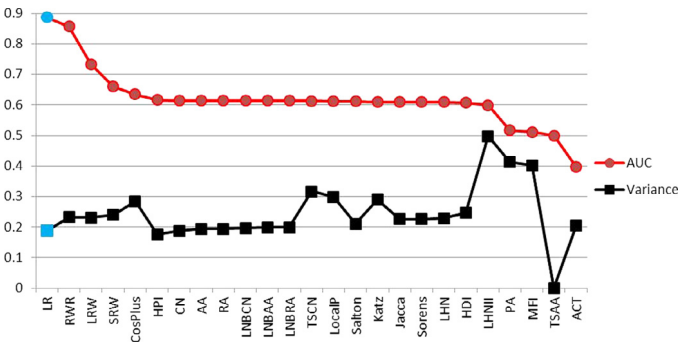

Fig. 8. Comparisons for the spurious link identification accuracy measured by *AUC*. Each data point is obtained by averaging over 1000 independent implementations, and the bold number emphasizes the highest value. The best parameters for some indexes are: *LR*(0.1), *Katz*(0.01), *LHNII*(0.9), *RWR*(0.95), *LRW*(5), *SRW*(5), *SimRank*(0.8).

Table 4
Comparisons for the spurious links identification accuracy measured by *AUC*. Each data point is obtained by averaging over 1000 independent implementations, and the bold number emphasizes the highest value. The best parameters for some indexes are: *LR*(0.1), *Katz*(0.01), *LHNII*(0.9), *RWR*(0.95), *LRW*(5), *SRW*(5), *SimRank*(0.8). The variance value should multiply e−2 actually.

Index	*LR*	*CN*	*Salton*	*Jacca*	*Soren*	*HPI*	*HDI*	*LHN*	*AA*
AUC	**0.8873**	0.6143	0.6116	0.6096	0.6096	0.6164	0.6071	0.6091	0.6143
Variance	0.1876	0.1869	0.2104	0.2254	0.2254	0.1758	0.246	0.2288	0.1935
Index	*RA*	*PA*	*LNBCN*	*LNBAA*	*LNBRA*	*LocalP*	*Katz*	*LHNII*	*ACT*
AUC	0.6142	0.5169	0.6137	0.6134	0.6134	0.6118	0.6099	0.5986	0.3966
Variance	0.1946	0.413	0.1962	0.1987	0.1986	0.2984	0.2895	0.4966	0.206
Index	*Cos+*	*RWR*	*LRW*	*SRW*	*MFI*	*TSCN*	*TSAA*	*SimRank*	
AUC	0.6345	0.8568	0.7335	0.66	0.511	0.6128	0.5	0.6706	
Variance	0.2847	0.2317	0.2314	0.2392	0.4027	0.317	0	0.2361	

3.4. Combined effectiveness with the nodes' types effect T

It is claimed that nodes' types effect T index defined in formula (4) could well quantify the effects of nodes' types on their linking behaviors. Here, we empirically proved that all the existent methods have been enhanced on the accuracy if combined with T index, both in the missing links prediction task and the spurious links identification task, and the average rises are 8.39% (Fig. 10) and 7.24% (Fig. 11) respectively.

3.5. Algorithm robustness

In the above tasks, the training set is assumed to be entirely clean, however, in real cases, the reliability of the observed network data could not always be guaranteed, there is always filled with noisy or incomplete information in them, and that is exactly why the link prediction is so important. As a result, when applied to solve application problems, the link prediction

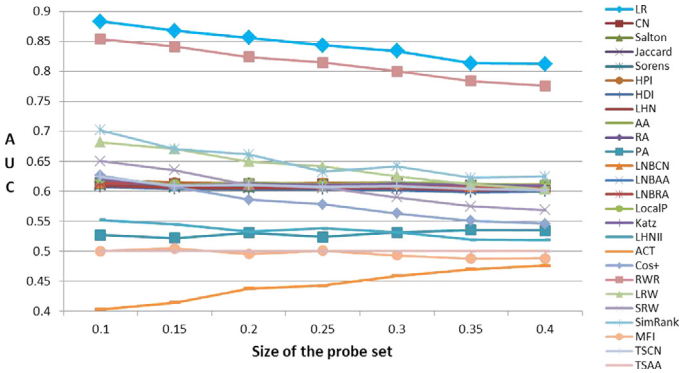

Fig. 9. Identifying spurious links for different sizes of probe set. Size of the probe set ranges from 0.1 to 0.4. Also each data point in the figure is averaged over 1000 independent implementations.

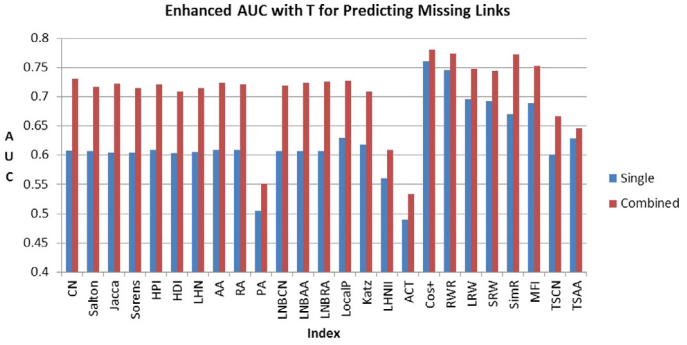

Fig. 10. Enhanced *AUC* with *T* for predicting missing links. *Ratio* of the training set is 0.9, and each data point is averaged over 1000 independent implementations. The blue bars are prediction accuracies obtained by the current methods, the red bars are prediction accuracies obtained by current methods combined with *T* index. The best parameters for some indexes are: *LR*(0.3), *Katz*(0.01), *LHNII*(0.9), *RWR*(0.95), *LRW*(5), *SRW*(5), *SimRank*(0.8). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

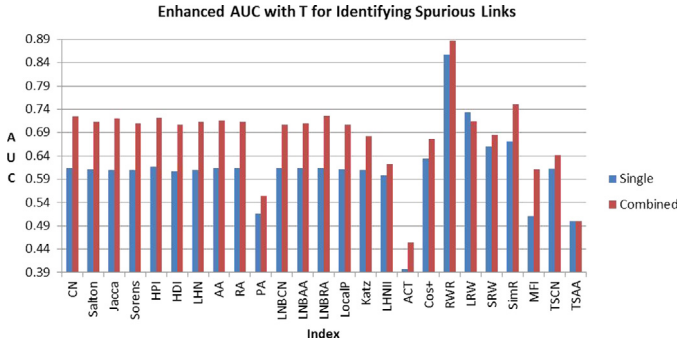

Fig. 11. Enhanced *AUC* with *T* for identifying spurious links. *Ratio* of the training set is 0.9, and each point is averaged over 1000 independent implementations. The blue bars are accuracies obtained by the current methods, the red bars are accuracies obtained by current methods combined with *T* index. The best parameters for some indexes are: *LR*(0.1), *Katz*(0.01), *LHNII*(0.9), *RWR*(0.95), *LRW*(5), *SRW*(5), *SimRank*(0.8). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

methods will most likely work under noisy environment. This section, we investigate to which extent each link prediction algorithm can resist the noise.

In the experiment, after the network is divided into the training set E^T and probe set E^P, some links are randomly added to or deleted from E^T, *ratio* is defined in Section 2.2.2, and to keep the training network connected, let *ratio* ranges from -0.3 to 1.

In Fig. 12. we show effects of random noise on link prediction results. More precisely, we investigate the dependence of *AUC* on $|ratio|$. Since it is hardly to distinguish each of them if putting on all the 26 indexes together, we just select 4 typical ones (*LNBRA* from local information based indexes, *LocalPath* from global information based indexes, *Cos+* from

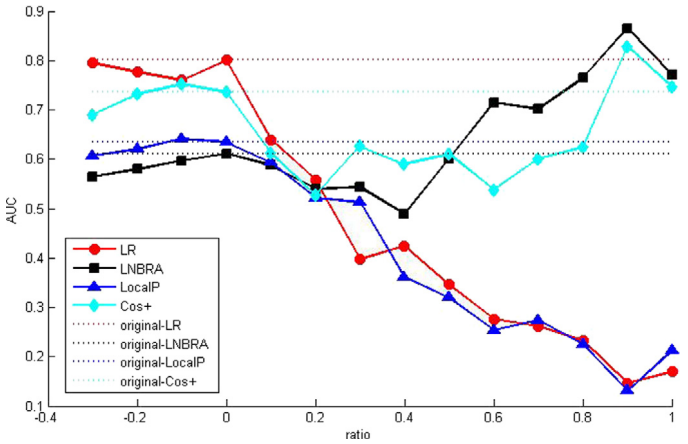

Fig. 12. The dependence of link prediction algorithms' accuracy (*AUC*) on *ratio* in the military network. The listed link prediction algorithm are *LR*, *LNBRA*, *LocalP*, *Cos+*. Dashed lines represent the *AUC* of these prediction algorithms in the clean network (i.e., the network without any noisy or missing links). $ratio < 0$ represents the missing link case and $ratio > 0$ stands for the noisy link case. Each data point is averaged over 100 independent implementations.

random walk based indexes, the left indexes nearly show similar curves as these four indexes), and compare them in Fig. 12, complete results see in S3.

It is observed that (1): *AUC* decreases with |*ratio*| generally for most indexes, and *AUC* of different link prediction algorithms decays with |*ratio*| with different speed. For example, *LR* has the highest *AUC* when $ratio = 0$, when $ratio = 100\%$ it reaches the lowest point among all. This indicates the performance of different link prediction methods in the military network may change dramatically when noise exists, and *R* value is such an index to quantify the different decay speed of *AUC*, which will be discussed later; (2) randomly removing links are less destructive than randomly adding link given the same |*ratio*| value. Taking the *Cos+* index as an example, adding 30% noisy links will decrease *AUC* from 0.7363 to 0.6260, the drop is 0.1103, while removing the same amount of random links will make the *AUC* be around 0.6906, the drop is 0.0457. This is on the contrary to results in Ref. [30], and the possible explanation may be that for the complex military organization, mistaken information (noisy links) may be more misleading (destructive) than missing information (missing links); (3) For *LNBRA* and *Cos+* methods, as the noise strength reaches a certain degree, adding noisy links may even improve the prediction accuracy, similar results were found in Ref. [31], which shows the recommendation accuracy can be improved by adding some virtual links. The hidden reason for this phenomenon may be that the random links improve the connectivity of network, while *LNBRA* and *Cos+* methods just work well on the network with high connectivity. However, it is not suitable for other methods, *AUC* of most indexes decrease rapidly as |*ratio*| increase (see in S3); (4) *LR* performs the best as the ratio ranges from -0.3 to 0.2, which means our method is the most robust and efficient one under the condition of small noise.

Next, we utilize *R* index defined in formula (11) to measure the algorithm robustness when links are randomly added or removed from the training set, and *R* depends on *ratio* apparently. In Fig. 13, we investigate the effects of *ratio* on *R*. And also we just show *LR*, *LNBRA*, *LocalPath* and *Cos+* for a clear illustration. Complete results see in S4.

It can be observed from Fig. 13 that: (1) the differences between algorithms' *R* values increase with *ratio* when *ratio* is positive, indicating the different decaying speed of different algorithms' robustness; (2) randomly adding spurious links is more destructive to the algorithms' robustness, take *LR* as an example, given the same |*ratio*| value, *R* is 1.056 when $ratio = -0.3$, while 0.8231 when $ratio = 0.3$; (3) *LR* maintains a comparative higher robustness under small noise environment ($-0, 1 \leq ratio \leq 0.1$). Fig. 14 demonstrates the average *R* values for all the indexes when *ratio* is between -0.1 and 0.1, and *LR* is ranked higher than most other prediction algorithms, indicating that it could better resist the small noise in the observed network than the majority of current methods.

4. Conclusion and discussion

This paper tries to address the problem of link prediction in complex military organization network, which is essential for commanders to obtain the qualified intelligence and make the right decisions. Three main contributions are made.

Firstly, we measured the nodes' types' effects on their linking behaviors based on the stochastic block model and proved its efficiency empirically.

Secondly, we proposed a new link prediction index for complex military organization network, considering both the nodes' types' effect and nodes' structural similarities, and proved it is remarkably superior to all the existent similarity-based indexes, both in predicting missing links and identifying spurious links.

Finally, we investigated link prediction algorithms' robustness in the military network under noisy environment, and found that: (1) spurious links are more destructive than missing links for prediction accuracy in military network, which is

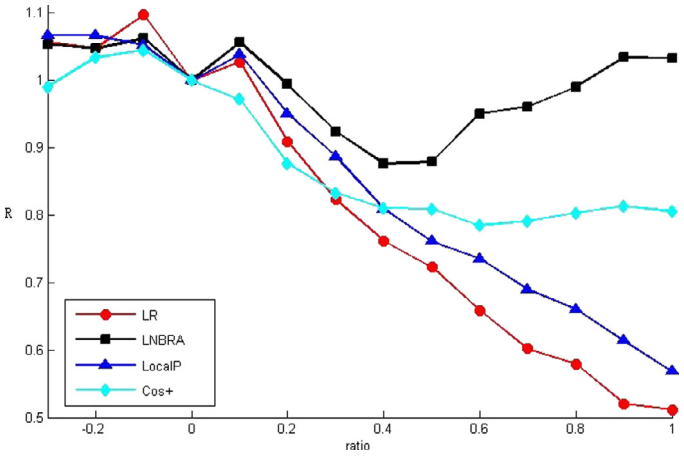

Fig. 13. The dependence of *R* on *ratio*. The listed link prediction algorithm are *LR*, *LNBRA*, *LocalP*, *Cos+*. *ratio* < 0 represents the missing link case and *ratio* > 0 stands for the noisy link case. Each data point is averaged over 100 independent implementations.

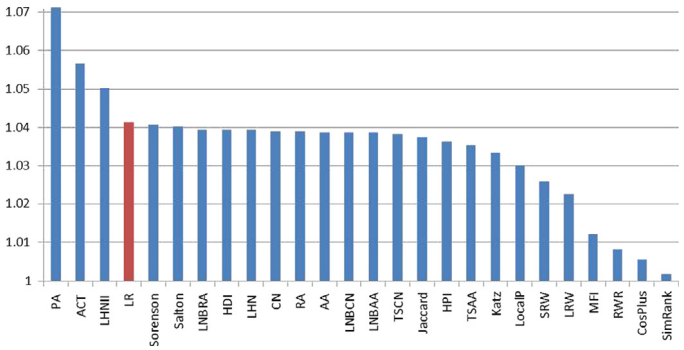

Fig. 14. Average *R* values of all the link prediction algorithms when *ratio* is between -0.1 and 0.1.

on the contrary to the results in Ref. [30], indicating that mistaken information may be more misleading in military areas; (2) for some indexes, such as *Cos+* and Local Naive Bayes based indexes, their prediction accuracies may even be improved with a certain spurious links existing, and the hidden reason may account for that the random spurious links improve the connectivity of network, and densely connected networks are easier predictable for these methods; (3) prediction accuracy and robustness are kind of contradictive, we have not found the index with both high accuracy and high robustness, some methods work well on clean networks but disable dramatically when noise exists, such as *RWR* index, and some methods perform ordinary on clean networks while are very robust that even some spurious noise may improve their accuracies, such as Local Naive Bayes based indices. Our method performs the best under noisy-free environment, and it still maintains the best accuracy under the condition of small noise, however, as the noise strength increases, its accuracy would be cut down sharply. All in all, *LR* is an efficient link prediction index for complex military organization network, with both high accuracy in predicting missing links and identifying spurious links and robustness under low noisy environment. Since the military area is very sensitive, the intelligence must be carefully artificially checked before we use the algorithm to purify it, which may reduce much obvious noise and leave a network with small latent noise, and that is just our method could still work well on, as a result, the method is practicable in reality.

Since our method is based on the FINC-E model, and FINC-E methodology is also applicable to criminal networks [32], disaster rescue [10], safety culture [33] and social management [34] and so on. Take the business organizations as an instance, the force nodes could be the sales forces and business markets; the intelligence nodes could be the market research and development institutes; and the C2 nodes could be the managers and decision-makers, thus our method could also be applied to these areas for many tasks, such as detecting the underground relationships between terrorists, predicting the potential business markets for decision-makers, and so on.

In the future, we should validate the method in more real military network data, and test its applications in the other social organization areas mentioned above. Besides, since the structure of the military organization is changeable overtime, link prediction analysis should be considered in the dynamic environment, as a result, we should further study the link prediction in time-dependent military networks.

Acknowledgments

The authors are grateful to the anonymous referees for their insightful suggestions and comments. This research was supported by the National Basic Research Program of China under Grant No. 71471176, No. 71471174, No. 61303266, No. 71301165 and No. 71522014. All authors acknowledge the National University of Defense Technology.

Appendix A. Proof of the Theorem 1

Proof. According to formula (3):

$$p(A_{xy}=1|A^o)=\frac{1}{Z}\sum_{P\in\Omega}\int_0^1|Q|p(A_{xy}=1|P,Q)p(A^o|A,Q)p(P,Q)\mathrm{d}Q$$

where $|Q|$ stands for the numbers of matrix elements, which are equal to the square of the number of groups in the network, and

$$Z=\sum_{P\in\Omega}|Q|p(A^o|P,Q)p(P,Q)\mathrm{d}Q \tag{A.1}$$

since $p(A_{xy}=1|P,Q)=Q_{\sigma_x,\sigma_y}$, then

$$p(A^o|P,Q)=\prod_{\alpha\le\beta}Q_{\alpha\beta}^{l^o_{\alpha\beta}}(1-Q_{\alpha\beta})^{r_{\alpha\beta}-l^o_{\alpha\beta}}. \tag{A.2}$$

Take formula (A.2) into formula (A.1), and get

$$Z=\sum_{P\in\Omega}\prod_{\alpha\le\beta}\int_0^1 Q_{\alpha\beta}^{l^o_{\alpha\beta}}(1-Q_{\alpha\beta})^{r_{\alpha\beta}-l^o_{\alpha\beta}}\mathrm{d}Q_{\alpha\beta}. \tag{A.3}$$

Let

$$H=\prod_{\alpha\le\beta}\int_0^1 Q_{\alpha\beta}^{l^o_{\alpha\beta}}(1-Q_{\alpha\beta})^{r_{\alpha\beta}-l^o_{\alpha\beta}}\mathrm{d}Q_{\alpha\beta}. \tag{A.4}$$

Next we prove:

$$H=\exp\left\{-\sum_{\alpha\le\beta}[\ln(r_{\alpha\beta}+1)+\ln\binom{r_{\alpha\beta}}{l^o_{\alpha\beta}}]\right\}. \tag{A.5}$$

Proof. With Beta integral formula:

$$\int_0^1 t^{a-1}(1-t)^{b-1}\mathrm{d}t=\frac{(a-1)!(b-1)!}{(a+b-1)!}$$

we have:

$$\int_0^1 Q_{\alpha\beta}^{l^o_{\alpha\beta}}(1-Q_{\alpha\beta})^{r_{\alpha\beta}-l^o_{\alpha\beta}}\mathrm{d}Q_{\alpha\beta}=\frac{l^o_{\alpha\beta}!(r_{\alpha\beta}-l^o_{\alpha\beta})!}{(r_{\alpha\beta}+1)!}=\frac{1}{r_{\alpha\beta}+1}\frac{l^o_{\alpha\beta}!(r_{\alpha\beta}-l^o_{\alpha\beta})!}{r_{\alpha\beta}!}$$

and

$$\ln H=\sum_{\alpha\le\beta}\ln\left(\frac{1}{r_{\alpha\beta}+1}\frac{l^o_{\alpha\beta}!(r_{\alpha\beta}-l^o_{\alpha\beta})!}{r_{\alpha\beta}!}\right)=-\sum_{\alpha\le\beta}[\ln(r_{\alpha\beta}+1)+\ln\binom{r_{\alpha\beta}}{l^o_{\alpha\beta}}].$$

Thus

$$H=\exp\left\{-\sum_{\alpha\le\beta}[\ln(r_{\alpha\beta}+1)+\ln\binom{r_{\alpha\beta}}{l^o_{\alpha\beta}}]\right\}. \quad \square$$

According to formula (A.5),

$$Z = \sum_{P\in\Omega} H = \sum_{P\in\Omega} \exp\left\{-\sum_{\alpha\le\beta}[\ln(r_{\alpha\beta}+1) + \ln(\binom{r_{\alpha\beta}}{l^o_{\alpha\beta}})]\right\} \tag{A.6}$$

and

$$\begin{aligned} p(A_{xy}=1|A^o) &= \frac{1}{Z}\sum_{P\in\Omega}\prod_{\alpha\le\beta}\int_0^1 p(A_{xy}=1|P,Q)Q_{\alpha\beta}^{l^o_{\alpha\beta}}(1-Q_{\alpha\beta})^{r_{\alpha\beta}-l^o_{\alpha\beta}}\mathrm{d}Q_{\alpha\beta} \\ &= \frac{1}{Z}\sum_{P\in\Omega}\prod_{\alpha\le\beta}\int_0^1 Q_{\sigma_x,\sigma_y}Q_{\alpha\beta}^{l^o_{\alpha\beta}}(1-Q_{\alpha\beta})^{r_{\alpha\beta}-l^o_{\alpha\beta}}\mathrm{d}Q_{\alpha\beta}. \end{aligned}$$

1^o. when $(\sigma_x,\sigma_y)\neq(\alpha,\beta)$,

$$\prod_{\alpha\le\beta}\int_0^1 Q_{\alpha\beta}^{l^o_{\alpha\beta}}(1-Q_{\alpha\beta})^{r_{\alpha\beta}-l^o_{\alpha\beta}}\mathrm{d}Q_{\alpha\beta} = \exp\left\{-\sum_{\alpha\le\beta,(\sigma_x,\sigma_y)\neq(\alpha,\beta)}\ln(r_{\alpha\beta}+1)+\ln(\binom{r_{\alpha\beta}}{l^o_{\alpha\beta}})\right\}.$$

2^o. when $(\sigma_x,\sigma_y)=(\alpha,\beta)$,

$$\begin{aligned} \int_0^1 Q_{\sigma_x,\sigma_y}Q_{\alpha\beta}^{l^o_{\alpha\beta}}(1-Q_{\alpha\beta})^{r_{\alpha\beta}-l^o_{\alpha\beta}}\mathrm{d}Q_{\alpha\beta} &= \int_0^1 Q_{\sigma_x\sigma_y}^{l^o_{\sigma_x\sigma_y}+1}(1-Q_{\sigma_x\sigma_y})^{r_{\sigma_x\sigma_y}-l^o_{\sigma_x\sigma_y}}\mathrm{d}Q_{\sigma_x\sigma_y} \\ &= \frac{(l^o_{\sigma_x\sigma_y}+1)!(r_{\sigma_x\sigma_y}-l)!}{(r_{\sigma_x\sigma_y}+2)!} \\ &= \frac{l^o_{\sigma_x\sigma_y}+1}{r_{\sigma_x\sigma_y}+2}\exp\{[\ln(r_{\alpha\beta}+1)+\ln(\binom{r_{\alpha\beta}}{l^o_{\alpha\beta}})]\}. \end{aligned}$$

Combine the above two situations, and reliability for link $\{v_x, v_y\}$ could be calculated as formula (A.7):

$$p(A_{xy}=1|A^o) = G_{xy} = \frac{1}{Z}\sum_{P\in\Omega}\frac{l^o_{\sigma_x\sigma_y}+1}{r_{\sigma_x\sigma_y}+2}H. \tag{A.7}$$

And the expression of Z, H see as formula (A.6) and formula (A.5).
For a specific TON, node type have been fixed, which means the division plans P are also fixed, then:

$$p(A_{xy}=1|A^o) = g_{xy} = \frac{l^o_{\sigma_x\sigma_y}+1}{r_{\sigma_x\sigma_y}+2}. \tag{A.8}$$

Theorem 1 has been proved. □

Appendix B. Supplementary data

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.physa.2016.11.097.

References

[1] R.E. Hayes, Measuring command and control (c2) effectiveness, in: MORS Workshop-Joint Framework for Measuring C2 Effectiveness, Citeseer, Laurel, MD, 2012.
[2] D.D. Šiljak, Large-Scale Dynamic Systems: Stability and Structure, Vol. 2, North Holland, 1978.
[3] D.S. Alberts, J.J. Garstka, R.E. Hayes, D.A. Signori, Understanding information age warfare, Tech. Rep., DTIC Document, 2001.
[4] D.S. Alberts, The Agility Advantage: A survival guide for complex enterprises and endeavors, DoD Command and Control Research Program, 2011.
[5] X. Hu, War complexity and war gaming & simulation in the information age, J. Syst. Simul. 18 (2006) 12.
[6] J. Cares, Distributed Networked Operations: The Foundations of Network Centric Warfare, iUniverse, 2006.
[7] K.M. Carley, D. Krackhardt, A typology for c2 measures, Tech. Rep., DTIC Document, 1999.
[8] K. Carley, D. Krackhardt, A pcans model of structure in organizations, in: Symposium on Command and Control Research and Technology, 1998, pp. 113–119.
[9] A.H. Dekker, C4isr architectures, social network analysis and the finc methodology: an experiment in military organisational structure, Tech. Rep., DTIC Document, 2002.
[10] A. Dekker, Applying social network analysis concepts to military c4isr architectures, Connections 24 (3) (2002) 93–103.
[11] A.H. Dekker, et al., C4isr, the finc methodology, and operations in urban terrain, J. Battlefield Technol. 8 (1) (2005) 25.
[12] A.H. Dekker, Network topology and military performance.
[13] G. Yang, W. Zhang, B. Xiu, Z. Liu, J. Huang, Key potential-oriented criticality analysis for complex military organization based on finc-e model, Comput. Math. Organ. Theory 20 (3) (2014) 278–301.
[14] L. Lü, T. Zhou, Link prediction in complex networks: A survey, Physica A 390 (6) (2011) 1150–1170.

[15] L. Lin-yuan, Link prediction on complex networks, J. Univ. Electron. Sci. Tech. China 39 (5) (2010) 651–661.
[16] L. Getoor, C.P. Diehl, Link mining: a survey, ACM SIGKDD Explor. Newslett. 7 (2) (2005) 3–12.
[17] R. Guimerà, M. Sales-Pardo, Missing and spurious interactions and the reconstruction of complex networks, Proc. Natl. Acad. Sci. 106 (52) (2009) 22073–22078.
[18] L. Lü, T. Zhou, Link Prediction, Higher Education Press, 2012.
[19] M. McPherson, L. Smith-Lovin, J.M. Cook, Birds of a feather: Homophily in social networks, Annu. Rev. Sociol. (2001) 415–444.
[20] A. Popescul, L.H. Ungar, Statistical relational learning for link prediction, in: IJCAI Workshop on Learning Statistical Models from Relational Data, Vol. 2003, Citeseer, 2003.
[21] B. Taskar, M.F. Wong, P. Abbeel, D. Koller, Link prediction in relational data, in: Proceedings of the Neural Information Processing Systems, 2003.
[22] Y. Yang, N. Chawla, Y. Sun, J. Hani, Predicting links in multi-relational and heterogeneous networks, in: Proceedings of the 2012 IEEE 12th International Conference on Data ining, 2012.
[23] R.N. Lichtenwalter, N.V. Chawla, Vertex collocation profiles: Subgraph counting for link analysis and prediction, in: Proceedings of the 21st International Conference on World Wide Web, 2012.
[24] D. Davis, R. Lichtenwalter, N.V. Chawla, Multi-relational link prediction in heterogeneous information networks, in: Proceedings of the 2011 International Conference on Advances in Social Networks Analysis and Mining, 2011.
[25] P.W. Holland, K.B. Laskey, S. Leinhardt, Stochastic blockmodels: First steps, Soc. Networks 5 (2) (1983) 109–137.
[26] S. Brin, L. Page, The anatomy of a large-scale hypertextual web search engine, 1998, in: Proceedings of the Seventh World Wide Web Conference, 2007.
[27] H. Tong, C. Faloutsos, J.-Y. Pan, Fast random walk with restart and its applications.
[28] M.-S. Shang, L. Lü, W. Zeng, Y.-C. Zhang, T. Zhou, Relevance is more significant than correlation: Information filtering on sparse data, Europhys. Lett. 88 (6) (2010) 68008.
[29] J.A. Hanley, B.J. McNeil, The meaning and use of the area under a receiver operating characteristic (roc) curve, Radiology 143 (1) (1982) 29–36.
[30] P. Zhang, X. Wang, F. Wang, A. Zeng, J. Xiao, Measuring the robustness of link prediction algorithms under noisy environment, Sci. Rep. 6 (2016).
[31] F. Zhang, A. Zeng, Improving information filtering via network manipulation, Europhys. Lett. 100 (5) (2012) 58005.
[32] C.E. Hutchins, M. Benham-Hutchins, Hiding in plain sight: criminal network analysis, Comput. Math. Organ. Theory 16 (1) (2010) 89–111.
[33] A. Sharpanskykh, S.H. Stroeve, An agent-based approach for structured modeling, analysis and improvement of safety culture, Comput. Math. Organ. Theory 17 (1) (2011) 77–117.
[34] M. Meyer, M.A. Zaggl, K.M. Carley, Measuring cmots intellectual structure and its development, Comput. Math. Organ. Theory 17 (1) (2011) 1–34.