LETTER

Measuring Collectiveness in Crowded Scenes via Link Prediction

Jun JIANG[†,††a)], **Di WU**[†], ***Members***, **Qizhi TENG**[†b)], **Xiaohai HE**[†], ***and*** **Mingliang GAO**[†††], ***Nonmembers***

SUMMARY Collective motion stems from the coordinated behaviors among individuals of crowds, and has attracted growing interest from the physics and computer vision communities. Collectiveness is a metric of the degree to which the state of crowd motion is ordered or synchronized. In this letter, we present a scheme to measure collectiveness via link prediction. Toward this aim, we propose a similarity index called superposed random walk with restarts (SRWR) and construct a novel collectiveness descriptor using the SRWR index and the Laplacian spectrum of a network. Experiments show that our approach gives promising results in real-world crowd scenes, and performs better than the state-of-the-art methods.

key words: *collectiveness, link prediction, random walk*

1. Introduction

In recent years, crowd analysis has garnered growing interest from the physics and computer vision communities, coupled with many potential applications such as intelligent video surveillance, crowd management, and smart virtual environments. Collective motion is one of the most common and fascinating phenomena in a large number of crowd systems [1]. Figure 1 gives some examples of this type of motion. Very recently, Zhou *et al.* [2] introduced the concept of collectiveness to describe the degree to which individuals act as one in realizing collective motion. In their work, crowd motion is modeled as a network, where the nodes represent the individuals, and each edge represents the correlation between two nodes. Then a collectiveness descriptor, denoted as Φ, is developed to compute the degree of the collective motion by utilizing the path similarities of the crowd motion network. The higher the value of Φ is, the more collective the crowd tends to be. Essentially, the notion of descriptor Φ is to recover all potential connections between the nodes of a network. Motivated by this, we regard measuring collectiveness as a link prediction problem [3].

Link prediction generally aims at estimating the presence probability of a link between the nodes of a network. In broad terms, the algorithms of link prediction fall into three categories: local similarity indices, global similarity indices, and quasi-local similarity indices [3]. The local indices predict a link score for a pair of nodes using only information regarding the immediate neighborhood of the node pair. The global indices compute a link score by taking into account the whole topological structure of the network. Actually, Zhou *et al.*'s method is equivalent to the KATZ index [4], which belongs to the global category. In contrast to the local indices, the global ones can provide more accurate prediction results. However, this type of index has two main drawbacks: 1) they depend heavily on the global topology of a network which is not easily obtained in many situations; 2) they incur high computational costs. To alleviate these problems and to achieve a good balance between accuracy and complexity, the quasi-local indices employ more information than local ones while dispensing with global topological knowledge.

Fig. 1 Examples of collective motion in crowds.

In this letter, we propose a scheme for measuring collectiveness in crowded scenes via link prediction. First, the crowd collective motion is modeled as a network. In contrast to [2], our method tries to find latent links in the network more accurately while reducing the time complexity. To this end, a novel quasi-local index called superposed random walk with restarts (SRWR) is constructed, and then applied to the network to uncover the individuals' connections. Finally, since the Laplacian spectrum encodes the structure and dynamics information about the network, it is incorporated to build a new collectiveness descriptor. Experiments validate the effectiveness and efficiency of the proposed approach, and demonstrate that it can achieve a better performance compared with the state-of-the-art methods [2], [5].

2. Measuring Crowd Collectiveness Based on Link Prediction

Let $G = (V, E, f)$ be an undirected simple weighted network, where V denotes a set of nodes, $|V| = N$, E denotes a set of links, and f is a weight function, $f : E \to R$, where R denotes real numbers. Here we merely consider nonnegative weights, namely, $f\left(e_{ij}\right) \geq 0$ for all $e_{ij} \in E$. Generally, G

Manuscript received January 14, 2015.
Manuscript publicized May 14, 2015.

[†]The authors are with the College of Electronics and Information Engineering, Sichuan University, Chengdu, China.
[††]The author is with the School of Computer Science, Southwest Petroleum University, Chengdu, China.
[†††]The author is with the School of Electrical & Electronic Engineering, Shandong University of Technology, Zibo, Shandong, China.
a) E-mail: scujiang@126.com
b) E-mail: nic5602@scu.edu.cn (Corresponding author)

DOI: 10.1587/transinf.2015EDL8011

can be represented by an $N \times N$ adjacency matrix $\mathrm{W} = \left[w_{ij}\right]$, whose entry $w_{ij} = f\left(e_{ij}\right)$ if there exists a link between node i and node j, and $w_{ij} = 0$ otherwise. In link prediction, a similarity score s_{ij}, $i, j \in V$, is assigned to every pair of nodes. In general, s_{ij} indicates the probability of a link between i and j, that is, the high score implies the link is more likely to exist.

2.1 Superposed Random Walk with Restarts

Inspired by the superposed random walk in [6] and random walk with restarts in [7], we derive a new index, i.e., SRWR. The strategy of this index is to use the random walk process in a quasi-local manner to estimate the similarity of node i with respect to node j.

First, the SRWR index normalizes the adjacency matrix W by column, that is, each of the columns of W adds up to 1. Clearly, w_{ij} represents the probability that a random walker staying at node j presently will move to node i in the next step. Given a random walker starting from node j, let $\mathbf{x}_j(t)$ be the $N \times 1$ state probability vector of this walker at step t, then we have

$$\mathbf{x}_j(t) = (1-c)\,\mathbf{y}_j + c\mathrm{W}\mathbf{x}_j(t-1), \tag{1}$$

where c is the probability of node j moving to a neighbor node, $1-c$ is the probability of restarting random walk from node j, and $\mathbf{y}_j$ is an $N \times 1$ vector with all elements equal to 0 other than the j-th element equals to 1. Initially, $\mathbf{x}_j(0) = \mathbf{y}_j$. Using the same policy in [6], we only take a few steps of random walk into account rather than the stationary state.

When a single random walker is released, the similarity between node i and node j is defined as

$$S_{ij}^{RWR}(t) = \mathbf{x}_{ij}(t) + \mathbf{x}_{ji}(t), \tag{2}$$

where $\mathbf{x}_{ij}(t)$ denotes the *i-th* element of $\mathbf{x}_j(t)$, and $\mathbf{x}_{ji}(t)$ has a similar meaning. After releasing multiple random walkers, we superpose the impacts of each random walker and get the similarity index

$$S_{ij}^{SRWR}(t) = \sum\nolimits_{m=1}^{t} S_{ij}^{RWR}(m). \tag{3}$$

2.2 Crowd Collectiveness Descriptor

The individuals' motion information is extracted from the tracklets [8] obtained by the KLT tracker [9], and then the fixed number of neighbors mechanism, that is, an individual interacts with K nearest neighbors [10], is employed to build a weighted network G' for modeling the crowd collective motion. The weight of each edge, w_{ij}, is determined by the velocity correlation between nodes i and j, namely,

$$w_{ij} = \max\left(\frac{\mathbf{v}_i \cdot \mathbf{v}_j}{\|\mathbf{v}_i\|\,\|\mathbf{v}_j\|}, 0\right), \tag{4}$$

where $\mathbf{v}_i$, $\mathbf{v}_j$ are the velocities of nodes i and j, respectively.

Then we apply the SRWR index to the network G' and obtain a similarity matrix S which reveals all pairs of individuals' motion consistency. Indeed, S can be viewed as the adjacency matrix of a new collective network G'' evolved from G'.

To measure the global crowd collectiveness, we calculate the Laplacian eigenvalues of G'' since the Laplacian spectrum can distill the structural and dynamical properties of a network. The Laplacian L on G'' is defined as

$$L_{ij} = \left(\sum_j S_{ij}\right)\delta_{ij} - S_{ij}, \tag{5}$$

where δ_{ij} is the Kronecker delta function [11]. After obtaining the eigenvalues of L: $\lambda_1, \lambda_2, \cdots, \lambda_N$, the collectiveness descriptor, denoted as Ψ, is defined as

$$\Psi = \sum\nolimits_{i=1}^{N} \lambda_i / N. \tag{6}$$

The framework of our crowd collectiveness descriptor algorithm is listed as follows.

Algorithm 1 Crowd collectiveness descriptor

Input: *The adjacency matrix W of the network G'*
Output: *The crowd collectiveness Ψ*
1. Apply the SRWR index to W, and get the network G''.
2. Compute the Laplacian spectrum of G''.
3. Ψ is the mean of the Laplacian eigenvalues.

The computational cost of Algorithm 1 consists mainly of two parts: the SRWR index and calculating eigenvalues of a matrix. The time complexity of SRWR with n steps is $O(N\langle k\rangle^n)$, where $\langle k\rangle$ is the average node degree of the network. Generally, the running time of eigenvalues algorithms is $O\left(N^3\right)$ [12]. In most cases, $\langle k\rangle \ll N$, thus the time complexity of our descriptor Ψ is approximately $O\left(N^3\right)$. For the descriptor Φ, the time complexity is the same as that of matrix inversion, which is $O\left(N^3\right)$. The method of [5] involves norm calculation and hence its time complexity is $O\left(N^2\right)$.

3. Experimental Results

We evaluate our method on the Collective Motion Database [2], which is composed of 413 video clips in crowed scenes. The ground truth is formed by asking ten persons to give a rate of the collective motion level, i.e., high, medium, and low, on each video clip. Furthermore, the high rate, the medium rate, and the low rate are mapped to numerical values 2, 1, and 0, respectively. Hence, the range of human-evaluated scores is [0, 20] when tallying up ten persons' votes. We use two metrics, classification accuracy and correlation, to quantify the performance of the proposed algorithm. More detailed descriptions about the database and the metrics can be found in [2]. The evaluation methodology on the Collective database is as follows: For a video clip, the collectiveness value is computed at each

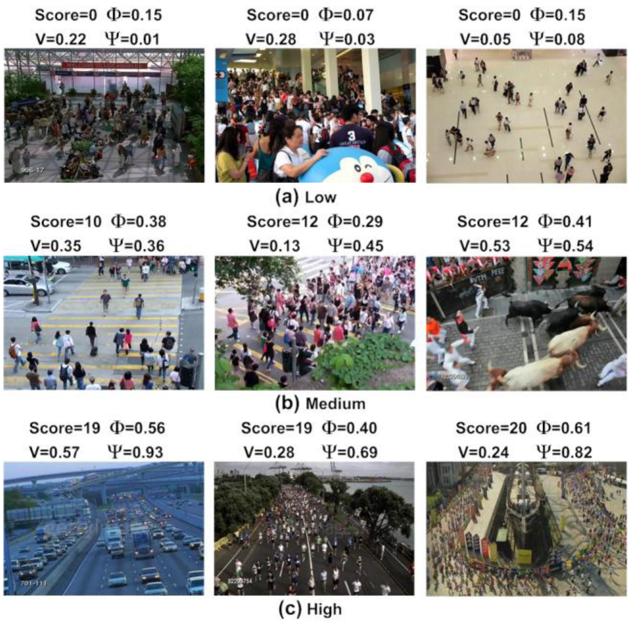

Fig. 2 Some typical examples of crowd collectiveness with their human-evaluated scores, Φ, **V** and Ψ in (a) low, (b) medium and (c) high category.

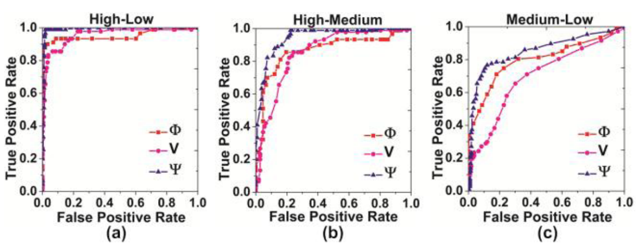

Fig. 3 ROC curves of three descriptors on various classification tasks: (a) high-low, (b) high-medium, and (c) medium-low.

frame, and the final collective score is determined by averaging over the total number of frames in the clip. The parameter c is empirically set to 0.85, and K is set to 20. The step t can be determined by the average shortest distance of a network. For simplicity, it is set to 3 in most of our experiments. We compare the performance of our descriptor with the state-of-the-art descriptors Φ [2] and **V** [5]. All descriptors are implemented in Matlab 2011 on a PC with a single Intel Core2 2.3 GHz processor and 2 GB RAM.

Some representative examples of the human-evaluated scores, Φ, **V** and Ψ for three collective motion levels are shown in Fig. 2. Clearly, Φ and Ψ can provide satisfactory collectiveness values for the low and medium categories, while **V** is prone to being unstable. For high category, Φ and **V** give relatively low values for some clips, whereas our Ψ can still offer reasonable values in these cases.

The ROC curves in Fig. 3 correspond to three binary classification tasks, i.e., high-low, high-medium, and medium-low, respectively. Table 1 lists the best accuracies. Obviously, Ψ achieves the highest accuracy, which is significantly better than Φ and **V**. The average computational time costs for a video clip of three algorithms are listed in

Table 1 The best accuracy comparisons of three descriptors

Classification	Descriptor	Best Accuracy (%)
High-Low	Φ	95
	V	93
	Ψ	**98**
High-Medium	Φ	84
	V	80
	Ψ	**88**
Medium-Low	Φ	79
	V	71
	Ψ	**85**

Table 2 Computational time costs of three descriptors

Descriptor	Time cost (secs)
Φ	148.8
V	**78.5**
Ψ	117.4

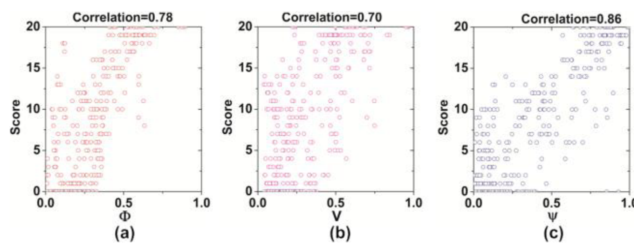

Fig. 4 Scatters of human-evaluated scores with respect to (a) Φ, (b) **V** and (c) Ψ.

Table 2, in which we compare only the collectiveness calculation phase of all descriptors while ignoring the motion feature extraction phase because its running time is the same for them. As can be seen, Ψ has the second fastest speed. Although both descriptors Φ and Ψ have the same time complexity, Ψ consumes less computational time compared to Φ because Ψ produces a symmetric and sparse similarity matrix which in turn reduces the running cost [13].

Figure 4 illustrates the correlations between the human-evaluated scores and Φ, **V**, and Ψ, respectively. In comparison with Φ and **V**, the proposed Ψ exhibits higher positive correlation which indicates it matches better with human perception.

4. Conclusion

In this letter, we have presented a link-prediction-based scheme for measuring collectiveness quantitatively in crowded scenes. We proposed a quasi-local SRWR index to predict missing links in a network and hence to uncover the correlations among individuals in the crowd. Furthermore, a novel collectiveness descriptor was constructed using the SRWR index and the Laplacian spectrum of the network. Experimental results verified the effectiveness and efficiency of the proposed method and demonstrated that it provided superior performance compared with the state-of-the-art methods.

Acknowledgments

The authors would like to sincerely thank the anonymous reviewers for their invaluable suggestions. This work was supported by NSAF under grant no. 11176018, National Natural Science Foundation of China under grant no. 61071161 and Promotive Research Fund for Young and Middle-aged Scientists of Shandong Province no. BS2014DX009.

References

[1] T. Vicseka and A. Zafeiris, "Collective motion," Phys. Rep., vol.517, no.3-4, pp.71–140, Aug. 2012.
[2] B.L. Zhou, X.O. Tang, and X.G. Wang, "Measuring crowd collectiveness," IEEE Conf. Computer Vision and Pattern Recognition (CVPR), pp.3049–3056, 2013.
[3] L.Y. Lü and T. Zhou, "Link prediction in complex networks: a survey," Physica A, vol.390, no.6, pp.1150–1170, March 2011.
[4] L. Katz, "A new status index derived from sociometric analysis," Psychometrika, vol.18, no.1, pp.39–43, March 1953.
[5] T. Vicsek, A. Czirók, E. Ben-Jacob, I. Cohen, and O. Shochet, "Novel Type of Phase Transition in a System of Self-Driven Particles," Phys. Rev. Lett., vol.75, no.6, pp.1226–1229, Aug. 1995.
[6] W.P. Liu and L.Y. Lü, "Link prediction based on local random walk," Europhysics Letters, vol.89, no.5, 58007, March 2010.
[7] J.-Y. Pan, H.-J. Yang, P. Duygulu, and C. Faloutsos, "Automatic multimedia cross-modal correlation discovery," Proceedings of the tenth ACM SIGKDD international conference on knowledge discovery and data mining (KDD), New York, USA, pp.653–658, 2004.
[8] W. Ge and R.T. Collins, "Multi-target Data Association by Tracklets with Unsupervised Parameter Estimation," Proceedings of the British Machine Conference, pp.93.1–93.10, Sept. 2008.
[9] C. Tomasi and T. Kanade, "Detection and tracking of point features," Int. J. Comput. Vision, vol.45, no.3, pp.1220–1234, 1991.
[10] N.W.F. Bode, D.W. Franks, and A.J. Wood, "Limited interactions in flocks: relating model simulations to empirical data," J. R. SOC. INTERFACE, vol.8, no.55, pp.301–304, Feb. 2011.
[11] P.N. McGraw and M. Menzinger, "Laplacian spectra as a diagnostic tool for network structure and dynamics," Phys. Rev. E, vol.77, no.3, 031102, March 2008.
[12] G.H. Golub and C.F. Van Loan, Matrix Computations (4th ed.), Johns Hopkins University Press, Maryland, 2012.
[13] D.R. Fokkema, G.L.G. Sleijpen, and H.A. Van der Vorst, "Jacobi–Davidson style QR and QZ algorithms for the reduction of matrix pencils," SIAM J. Sci. Comput., vol.20, no.1, pp.94–125, Aug. 1998.