

基于图神经网络的机会网络节点重要度评估方法

刘琳岚¹ 谭镇阳¹ 舒 坚²

¹(南昌航空大学信息工程学院 南昌 330063)

²(南昌航空大学软件学院 南昌 330063)

(liulinlan@nchu.edu.cn)

Node Importance Estimation Method for Opportunistic Network Based on Graph Neural Networks

Liu Linlan¹, Tan Zhenyang¹, and Shu Jian²

¹(School of Information Engineering, Nanchang Hangkong University, Nanchang 330063)

²(School of Software, Nanchang Hangkong University, Nanchang 330063)

Abstract Opportunistic network is a type of self-organized networks which uses the opportunity of a node moving to realize communication. Because of opportunistic communication mode, opportunistic network has observable time-varying and dynamic characteristics. The estimation of node importance is the key to study the information dissemination of opportunistic network. A novel node importance estimation method based on graph neural network (GNN-NIE) framework is proposed. Opportunistic network is sliced into opportunistic network units which is modeled by aggregate graph to present network information. The dynamic network embedding model is employed to extract the temporal and structural information among the opportunistic network units, so as to obtain the dynamic attribute features of each node in the network. Taking advantage of the GNN's ability of extracting the features of graph data, the relationship between node dynamic attribute features and the node importance is achieved, so that the node importance of opportunistic network is estimated. The results on three real opportunistic network datasets MIT reality, Haggle project and Asturias-er show that compared with the temporal degree, temporal betweenness, temporal PageRank, and kshell-CN, the proposed method has faster propagation rate, larger message coverage and better SIR and NDCG@10 values.

Key words opportunistic network; node importance; network embedding; graph neural network; message propagation rate

摘要 机会网络(opportunistic network)是一种利用节点移动的相遇机会实现通信的自组织网络,机会式的通信方式导致其具有时变性与动态性,节点重要度的评估是研究机会网络信息传播的关键.提出一种基于图神经网络的机会网络节点重要度评估方法.将机会网络进行时间切片,对得到的机会网络单元采用聚合图建模,以表征网络信息;采用动态网络嵌入模型提取机会网络单元间的时序变化信息、拓扑结构信息,得到网络的动态属性特征;借助图神经网络(graph neural network, GNN)在图数据处理上的优势,获得网络动态属性特征与节点重要度之间的映射关系,实现节点重要度的评估.在3个真实机会网络数据集MIT,Haggle,Asturias-er上的实验结果表明:相比于时效介数(temporal betweenness,

收稿日期:2020-08-31;修回日期:2021-03-29

基金项目:国家自然科学基金项目(62062050,61962037);江西省自然科学基金项目(20202BABL202039)

This work was supported by the National Natural Science Foundation of China (62062050, 61962037) and the Natural Science Foundation of Jiangxi Province (20202BABL202039).

通信作者:舒坚(shujian@nchu.edu.cn)

TB)方法、时效度(temporal degree, TD)方法、时效 PageRank(temporal PageRank 和 t-PageRank)方法以及 kshell-CN 方法,该方法具有更快的消息传播速率和更大的消息覆盖范围,其 SIR 和 NDCG@10 指标更优。

关键词 机会网络;节点重要度;网络嵌入;图神经网络;消息传播速率

中图法分类号 TP393

机会网络(opportunistic network, ON)^[1]是一种不需要通信双方的节点之间存在完整链路,利用节点移动带来的间歇性连通机会实现通信的自组织动态网络。机会网络的部分概念源于早期的延迟容忍网络(delay tolerant network, DTN)研究,其主要的目标之一就是为了支持那些具有间歇性的连通、延时量大、错误率高等特点的异构网络之间的互联与相关,如车载自组织网、移动数据分流、信息分享、移动计算等。机会网络拓展了网络通信的使用范围,将成为未来泛在通信的重要组成部分^[2]。

机会网络利用机会式的通信方式实现智能设备间的内容、资源和服务的共享。正是这种机会式的通

信方式导致了机会网络具有明显的动态性与时变性。因此,如何选择最理想的转发目标以及最理想的转发时机成为机会网络中数据传输机制关注的核心问题^[3]。对节点重要度的评估能够辅助上层路由协议选取最佳转发目标,从而提高整网的消息投递率。

一个具有多个移动节点的机会网络中,源节点 A 向目标节点 B 传递消息的过程如图 1 所示。由于机会网络节点间通信的频繁连接和断开,且移动节点的电池能量、内存等资源相对受限,为了使源节点 A 的时效性消息能够有效地投递至目标节点 B,应选择重要度(即消息传播能力)高的节点作为转发目标,从而提高消息的投递成功率。

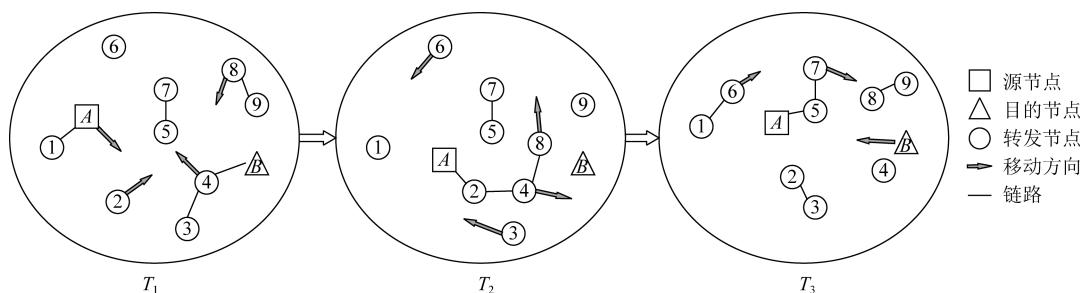


Fig. 1 Schematic diagram of opportunistic network communication

图 1 机会网络通信示意图

针对机会网络中通信节点位置的不确定性以及节点对连接的周期性导致其网络结构随时间频繁变化的特点,本文采用动态网络嵌入方法将随时间变化的动态网络拓扑信息压缩到特征空间中,以获取网络动态属性特征,利用图神经网络(graph neural network, GNN)建立网络动态属性特征与节点重要度的映射关系,实现对节点重要度的评估。本文的主要贡献有 3 个方面:

- 1) 利用节点间的交互信息,通过聚合图模型对机会网络进行建模。针对机会网络动态性与时变性的特点,将时序上连续的机会网络按照时间片切分为离散的机会网络单元,在机会网络单元上建立聚合图模型,确定图模型中连边的权重以构建网络属性矩阵,表征机会网络的拓扑结构信息。
- 2) 提出一种适用于机会网络的动态网络嵌入

方法。采用类增量学习的方式,根据输入的网络属性矩阵生成节点的嵌入表示。提取机会网络的拓扑结构信息与时序变化信息,将其压缩表示为由节点动态属性嵌入向量组成的网络动态属性特征矩阵。

3) 提出基于图神经网络的机会网络节点重要度评估模型。通过节点之间相互关系对节点动态属性嵌入向量进行更新,实现节点邻域信息的融合,建立节点动态属性嵌入向量与节点重要度之间的映射关系,评估机会网络节点的重要度。

1 相关研究

近年来,评估节点重要度的方法可分为 3 类:基于局部指标的评估、基于全局指标的评估和基于混合指标的评估。

1.1 基于局部结构的评估方法

网络的拓扑结构在很大程度上能够决定一个节点在网络中的重要度。因此,依据网络的拓扑结构设计节点的重要度排序指标,是一类对网络节点重要度进行评估的重要方法。节点的重要度也称为“中心性”,即网络中节点的重要程度取决于该节点与其他节点的连接使其具有的显著性^[4]。度中心性^[4](degree centrality, DC)的基本思想可描述为网络中节点的邻居数量决定其节点本身的影响力大小。文献[5]采用路径流的方式对时序网络建模,提出时序度中心性,将静态网络中的度中心性指标拓展到时序网络,考虑了节点度随时间变化的波动,能够有效评估动态网络中的节点重要度。为进一步提高节点重要度评估的准确性,Ren 等人^[4]提出一种利用节点在网络中半局部信息的节点重要度评估指标,简称半局部中心性(semi-local centrality, SLC),该指标通过计算网络中节点的 2 层邻居数量来表征节点的重要度,在损耗较少计算时间的前提下得到了较好的评估效果。度中心性与半局部中心性利用节点在网络拓扑中的领域信息对节点的重要度进行评估,认为邻域内节点数目相同的节点其重要程度也相同。

1.2 基于全局结构的评估方法

在现实世界中的诸多网络中,存在着一些邻居数目很少但重要程度却不容小觑的节点,这类节点的重要性往往不能采用传统的局部结构评估方法度量。因此,基于全局结构的评估方法应运而生。接近中心性(closeness centrality, CC)^[4]指标通过计算节点与网络中其他所有节点的平均距离来衡量该节点在网络全局结构上的影响力,从而消除特殊位置对的节点重要度评估的干扰。

Katz 中心性^[4]指标在接近中心性的基础上进行改进,不仅考虑了节点对之间的最短路径,还考虑了它们之间的其他非最短路径。介数中心性(betweenness centrality, BC)^[4]刻画了节点对在网络中沿最短路径传输的网络流的控制力。文献[6]基于静态网络中的 CC 以及 BC,提出了时效介数(temporal betweenness, TB)中心性指标评估时效网络中节点重要程度,TB 不仅考虑了经过节点的最短路径的数目,还考虑了最短路径上节点对于消息的保持时间,从而综合计算得到时序介数中心性的值,对节点重要度进行评估。

针对局部性指标仅仅只考察节点的局部特性而忽略节点在网络中位置的问题,Kitsak 等人^[7]提出 k -壳分解法(k -shell decomposition),该方法首先利

用节点位于网络中的位置,将外层节点层层剥去,认为处于内层的节点其重要程度也越高。针对 k -壳粒度粗的问题,学者们提出了一系列关于 k -壳分解法的改进方法,如将节点进行预分类的层次 k -壳分解法(hierarchical k -shell decomposition)^[8]、使用迭代因子进一步细化分解粒度的迭代因子 k -壳分解法(k -shell iteration factor)^[9]等。基于全局结构的评估方法能够考虑网络的结构信息,对网络中的节点重要性评判更加准确,但基于全局结构的方法时间复杂度较高,在大型网络和动态网络中的应用受到限制。

1.3 基于混合结构的评估方法

研究人员结合前 2 类指标的优点提出了基于混合指标的节点重要度评估方法。Zareie 等人^[10]结合节点的邻居信息、影响力范围以及范围内的影响力强度提出了多样性强度中心性(diversity strength centrality, DSC)指标对节点的重要度做出评估。文献[11]综合考虑节点的拓扑结构、动力学特征,采用时序游走得到节点的路径信息,并定义了传播中心性和接收沟通性指标。文献[12]对传统的 h-index 指标进行改进,考虑了传统 h-index 未考虑的全局性信息,提出了扩展的 h-index(exended h-index, EH)指标。文献[13]提出一种快速的混合指标评估方法,通过计算网络中节点的简单指标(DC、聚类系数、 k -shell)与复杂指标(CC、BC、网络效率)构建知识库,利用层次分析法将复杂指标进行融合得到重要度指标,再通过最小二乘支持向量机找到简单指标与重要度指标的映射关系。文献[14]采用 4 种传统的中心性指标对多层网络中不同网络层节点的中心性和权威性进行量化,得到一个 4 维张量,通过张量计算将网络中所有节点和层的信息结合起来,得到张量奇异向量中心性指标评估多层网络中节点的重要性。

近年来,研究人员从数据挖掘的角度,利用机器学习算法对网络中的节点重要度进行评估。文献[15]利用网络嵌入技术提取电力网络的网络信息,结合节点本身的属性信息,利用支持向量回归机对电力网络的节点重要度进行评估。

上述研究为解决机会网络节点重要度评估问题提供了丰富的思路和解决方案,但多数方法是针对拓扑结构变化缓慢的社交网络所设计^[16-21],无法直接应用于机会网络。本文针对机会网络的动态性与时变性特点,对机会网络进行切片处理,将切片后的网络快照表示为机会网络单元,通过对机会网络单元建立图模型,确定图模型中节点间连边的权值构建网络属性矩阵,表征网络拓扑信息;采用基于自编

码器结构的动态网络嵌入方法,将网络属性矩阵作为输入,从网络的时序变化信息以及拓扑结构信息 2 个方面获取机会网络的网络动态属性特征;引入图神经网络构建节点重要度评估模型,实现节点间邻域信息的融合,从而建立网络动态属性特征与节点重要度的映射关系,对机会网络节点的重要度进行评估。

2 网络拓扑信息的表征

本文将连续的机会网络数据按照时间切片划分为离散的机会网络单元,建立机会网络单元对应的图模型以构建机会网络单元的带权邻接矩阵(本文称网络属性矩阵),将多个网络属性矩阵组成的张量作为动态网络嵌入模型的输入,以进一步提取网络信息。

2.1 机会网络数据的转化

本文主要研究机会网络中的便携设备交换网络(pocket switched network, PSN),其数据集的数据格式如图 2 所示:

连接时间	连接状态	源节点	目标节点	连接类型
------	------	-----	------	------

Fig. 2 Data format

图 2 数据格式

利用原始数据中节点对的连接情况,计算节点对之间的连接持续时长以及连接次数。具体地,节点对[1,3]在第 1 秒连接类型为“up”,在第 3 秒的连接类型为“down”,则连接时长为 $3 - 1 = 2$ s,节点对[1,3]在整个原始数据中的 $\langle \text{up}, \text{down} \rangle$ 状态成对存在的次数为 12 次,则节点对[1,3]的连接次数为 12 次。

2.2 机会网络单元的切分

根据机会网络的动态性与时变性,将连续的原始数据集按照时间片切分为离散的机会网络单元。原始的机会网络表示为 G ,按照时间片进行切分可将连续的机会网络表示为由一系列机会网络单元组成的有序集合 $G = \{G_1, G_2, \dots, G_N\}$ 。其中, $N = T/\Delta t$ 为机会网络单元的数量, T 为整个机会网络数据集持续时长, Δt 为机会网络单元切片时长, G_1 表示初始机会网络单元, G_2 表示间隔一个切片时长的机会网络单元,依此类推, G_N 为最后一个机会网络单元,以 MIT 数据集为例,将其网络拓扑结构可视化后,其演化过程如图 3 所示。对机会网络单元建立对应的聚合图模型,使用 $G_t = \{V_t, E_t, W_t\}$ 表示第 t ($t = 1, 2, \dots, N$) 个机会网络单元的聚合图, V_t 表示在第 t 个机会网络单元上的节点集合, E_t 表示在第 t 个机会网络单元上边的集合, W_t 表示第 t 个机会网络单元上对应边上的权值的集合。

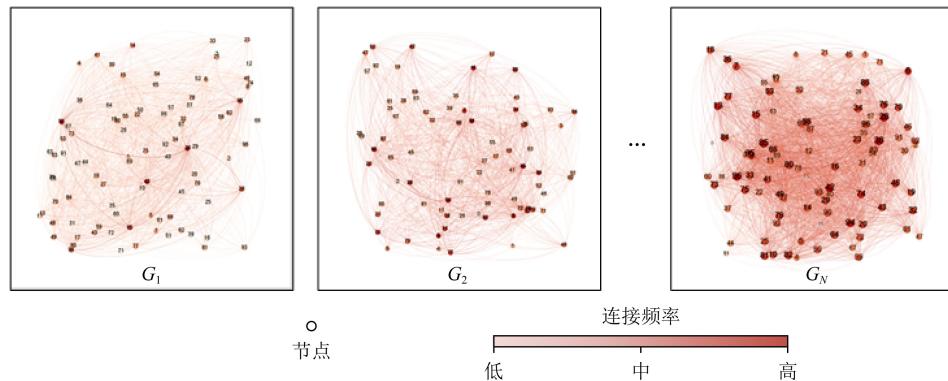


Fig. 3 Evolution process of opportunistic network topology

图 3 机会网络拓扑的演化过程

机会网络单元切片时长 Δt 的长短直接关系到网络拓扑信息的表征能力。针对该问题,文献[22]对具有代表性的机会网络数据集进行统计分析,从机会网络的持续性与周期性 2 个角度,提出了基于邻接相关系数、聚类系数和节点平均度确定切片时长的方法。借鉴该方法,一方面考虑到切片时长 Δt 太短导致机会网络单元拓扑图过于稀疏,无法准确描述机会网络中节点间的连接情况,另一方面考虑切

片时长 Δt 太长导致机会网络单元内拓扑图过于密集、节点间连接频繁,导致无法准确区分节点重要度的情况。本文对所使用的 3 个数据集按照不同切片时长进行切片得到机会网络单元,分别统计每个机会网络单元内有效记录的次数,其结果如图 4~6 所示。

按照 $\Delta t \in \{0.5 \text{ d}, 1 \text{ d}, 10 \text{ d}, 30 \text{ d}\}$ 对 MIT 数据集进行切片,分别统计每个机会网络单元内的有效记录情况。由图 4 可知,当 $\Delta t = 0.5 \text{ d}$ 和 $\Delta t = 1 \text{ d}$ 时,有效

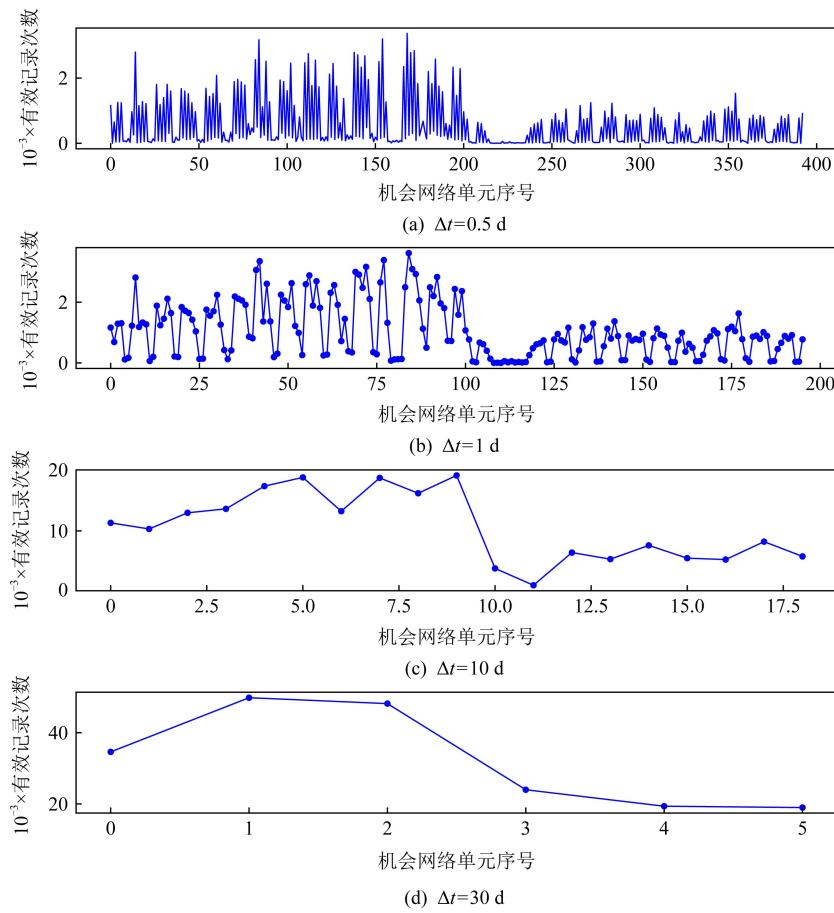


Fig. 4 The effective records in each opportunistic network unit of MIT under different slice time length

图 4 不同切片时长下 MIT 各机会网络单元内有效记录

记录次数呈现明显的波动性,大多数机会网络单元内的有效记录次数非常少,导致网络拓扑结构过于稀疏,节点大多处于未连接状态,因此无法获取有效的拓扑信息.当 $\Delta t = 30$ d 时,有效记录次数均大于 20 000 条,大部分节点间都发生频繁的连接,导致网络拓扑过于密集,无法通过拓扑结构信息区分节点的重要度.因此,采用 $\Delta t = 10$ d 作为 MIT 数据集的候选切片时长,通过实验确定最终切片时长.

按照 $\Delta t \in \{0.5\text{ h}, 1\text{ h}, 4\text{ h}, 8\text{ h}\}$ 对 Haggle 数据集进行切分统计各机会网络单元内的有效记录情况.由图 5 可知,当切片时长 $\Delta t = 0.5$ h 与 $\Delta t = 1$ h 时,有效记录数呈明显的波动性,且大多数机会网络单元内不存在有效的连接,导致机会网络单元形成的网络拓扑中绝大部分节点处于孤立状态,网络拓扑极为稀疏,无法通过网络的拓扑连接情况对节点的重要度进行评估.当切片时长为 $\Delta t = 8$ h 时,有效连接数量的均值大于 25 000 条,大部分节点频繁连接,导致网络拓扑过于密集,无法通过拓扑结构信息

对不同节点的重要度做出评估.因此,将有效记录数量过于稀疏或密集的时间切片长度剔除,采用 $\Delta t = 4$ h 作为 Haggle 的候选切片时长,通过实验确定最终切片时长.

按照 $\Delta t \in \{6\text{ h}, 24\text{ h}, 48\text{ h}, 96\text{ h}\}$ 对 Asturias-er 数据集进行切片,分别统计每个机会网络单元内的有效记录情况.由图 6 可知,当 $\Delta t = 6$ h 和 $\Delta t = 24$ h 时,得到的有效记录次数呈现明显的波动性,机会网络单元内大多数机会网络单元网络拓扑结构过于稀疏,节点大多属于未连接状态,因此无法获取有效的拓扑信息.当 $\Delta t = 96$ h,机会网络单元内的有效记录次数均大于 60 000 条,节点间连接频繁,无法通过拓扑结构信息区分节点的重要度.因此,采用 $\Delta t = 48$ h 作为 Asturias-er 数据集的候选切片时长,通过实验确定最终切片时长.

2.3 网络属性矩阵的构建

按照切片时长将连续的机会网络切分为离散的

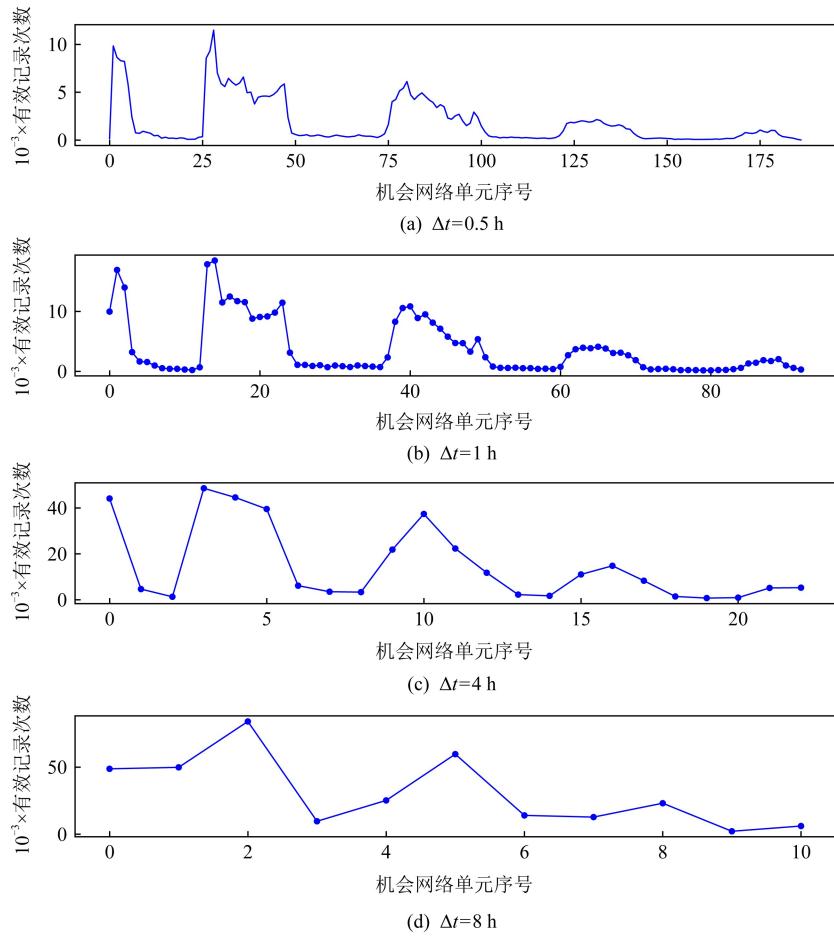


Fig. 5 The effective records in each opportunistic network unit of Haggle under different slice time length

图 5 不同切片时长下 Haggle 各机会网络单元内有效记录

机会网络单元并建立相应的图模型, 考虑到机会网络中的通信是以人携带智能手机、蓝牙通信设备进行数据交换, 人具有社会性, 相互通信的持续时长的频繁程度取决于通信双方的亲密程度。因此, 为了准确描述机会网络中节点间的关系, 本文采用节点间的连接次数以及连接持续时长表征节点间连边的权重。具体地, 在机会网络单元 $G_t = (V_t, E_t, W_t)$ 中边的权值 W_t 可表示为该机会网络单元内存在连接的节点间的连接次数与连接持续时长之积

$$W_t = \{w_{ij}^t \mid \forall (i \neq j) \cap (i, j) \in E_t\}, \quad (1)$$

其中, $w_{ij}^t = c^t \sum_{k=1}^q d_k^t$, c^t 为节点对 (i, j) 在第 t 个机会网络单元上的连接次数, d_k^t 为节点对 (i, j) 在第 t 个机会网络单元上第 k 次连接的持续时长, q 为节点对 (i, j) 在第 t 个机会网络单元上的连接总次数。

依据机会网络单元的拓扑状态来构建相应的带权邻接矩阵(本文称为网络属性矩阵)为

$$\mathbf{M}_t = (\mathbf{m}_1^t, \mathbf{m}_2^t, \dots, \mathbf{m}_n^t)^T = \begin{pmatrix} 0 & w_{12}^t & \cdots & w_{1n}^t \\ w_{21}^t & 0 & \cdots & w_{2n}^t \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1}^t & w_{n2}^t & \cdots & 0 \end{pmatrix}, \quad (2)$$

其中, \mathbf{M}_t 为第 t 个机会网络单元对应的网络属性矩阵, $\mathbf{m}_i^t = (w_{1i}^t, w_{2i}^t, \dots, w_{ni}^t)$ 是由 \mathbf{M}_t 的第 i ($i = 1, 2, \dots, n$) 行元素构成的 n 维向量, 即节点 i 的属性向量, w_{uv}^t 为节点对 u 和 v 间连边的权值($u, v = 1, 2, \dots, n$), n 为机会网络节点总数。

3 节点重要度评估

本文以网络属性矩阵作为动态网络嵌入模型的输入, 利用动态网络嵌入方法将机会网络单元之间的时序变化信息以及拓扑结构信息压缩到特征空间中以获取网络动态属性特征, 并以节点动态属性特征

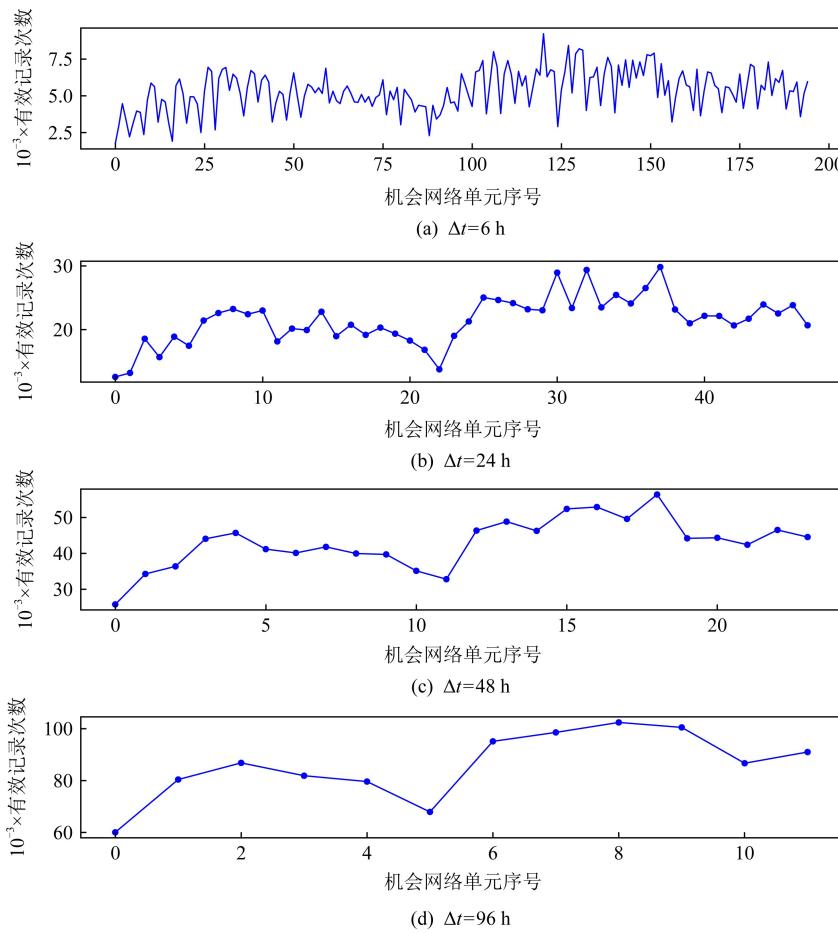


Fig. 6 The effective records in each opportunistic network unit of Asturias-er under different slice time length

图 6 不同切片时长下 Asturias-er 各机会网络单元内有效记录

向量的形式表示。之后,利用图神经网络构建节点重要度评估模型,建立节点动态属性特征与节点重要度的映射关系,从而对节点重要度进行评估。

3.1 动态网络嵌入模型的构建

本文采用基于自编码器的动态网络嵌入模型提取网络动态属性特征。本节在介绍模型结构的基础上,从时序信息的提取和拓扑结构信息的提取阐述模型的实现。

3.1.1 模型结构

自编码器模型可以通过一系列非线性映射操作将网络的原始特征映射到一个低维的特征空间中。模型的目的是尽可能地使得原始特征和重构特征保持相似性,从而达到在降维和嵌入过程中降低网络信息的损失。基于此,本文采用基于自编码器的动态网络嵌入模型提取网络动态属性特征,将机会网络对应的多个网络属性矩阵作为输入,每个网络属性矩阵都要经过一系列的解码和编码操作,同时为了

提取机会网络的时序变化信息以及拓扑结构信息,增加了编码阶段对原始特征整合的隐藏层和解码阶段对嵌入特征分解的隐藏层,最终得到由节点动态属性嵌入向量组成的网络动态属性嵌入矩阵作为模型的输出。动态网络嵌入模型的结构如图 7 所示。

图 7 中, G_t 为第 t 个机会网络单元, M_t 为第 t 个机会网络单元拓扑图对应的网络属性矩阵, $m_i^t \in \mathbb{R}^n$ 为第 t 个机会网络单元中节点 i 的属性向量, θ_{t-1} 与 θ_t 分别代表了在第 $t-1$ 个与 t 机会网络单元对应的自编码器的可训练的参数, x_i, \hat{x}_i 为自编码器模型的输入数据和输出的重构数据, $y_i^{(l)}, \hat{y}_i^{(l)}$ 表示节点 i 在第 l 个编码层和解码层的隐藏特征, $h_i \in \mathbb{R}^F$ 表示节点 i 编码器的输出,即节点 i 的动态属性嵌入向量。

模型包括编码与解码 2 部分:1)编码部分的作用是将原始特征向量映射到嵌入空间中;2)解码部分恢复嵌入向量到重构空间中。

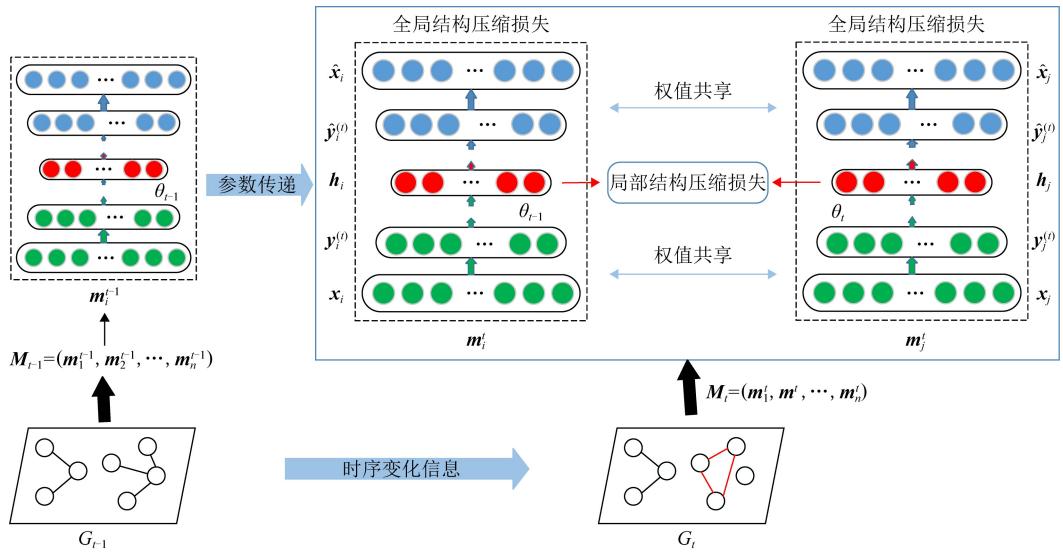


Fig. 7 Dynamic network embedding model based on autoencoder

图 7 基于自编码器的动态网络嵌入模型

在编码部分, x_i 代表节点的原始特征, 即机会网络单元中节点 i 的属性向量 m_i , y_i 为编码器部分隐藏层的潜在特征, 在嵌入空间中的编码结果表示为 $h_i \in \mathbb{R}^F$ (F 为节点嵌入维数), 这些变量之间的关系为

$$y_i^{(0)} = \sigma(\mathbf{W}^{(0)} x_i + \mathbf{b}_i^{(0)}), \quad (3)$$

$$y_i^{(l)} = \sigma(\mathbf{W}^{(l)} y_i^{(l-1)} + \mathbf{b}_i^{(l)}), l=1,2,\dots,L, \quad (4)$$

$$h_i = \sigma(\mathbf{W}^{(L+1)} y_i^{(L)} + \mathbf{b}_i^{(L+1)}), \quad (5)$$

其中, L 表示编码器隐层的总层数, $\mathbf{W}^{(l)}$ 为编码层第 l 层对应的权重矩阵, $\mathbf{b}_i^{(l)}$ 为编码层第 l 层对应的偏置矩阵, $\sigma(\cdot)$ 为激活函数。

在解码部分, 输入为动态属性嵌入向量 $h_i \in \mathbb{R}^F$, 输出为重构向量 \hat{x}_i , \hat{y}_i 表示隐藏层的潜在特征, 这些变量间的关系为

$$\hat{y}_i^{(L')} = \sigma(\hat{\mathbf{W}}^{(L'+1)} h_i + \hat{\mathbf{b}}_i^{(L'+1)}), \quad (6)$$

$$\hat{y}_i^{(l-1)} = \sigma(\hat{\mathbf{W}}^{(l)} \hat{y}_i^{(l)} + \hat{\mathbf{b}}_i^{(l)}), l=1,2,\dots,L', \quad (7)$$

$$\hat{x}_i = \sigma(\hat{\mathbf{W}}^{(0)} \hat{y}_i^{(0)} + \hat{\mathbf{b}}_i^{(0)}), \quad (8)$$

其中, L' 表示解码器隐层的总层数, $\hat{\mathbf{W}}^{(l)}$ 为解码层第 l 层对应的权重矩阵, $\hat{\mathbf{b}}_i^{(l)}$ 为解码层第 l 层对应的偏置矩阵, $\sigma(\cdot)$ 为激活函数。此外, 为了保证原始特征与重构特征的一致性, 本文中编码层隐层数 L 与解码器隐层数 L' 均设置为 4 层。

最终, 将机会网络对应的多个网络属性矩阵 $\mathbf{M} = (\mathbf{M}_1, \mathbf{M}_2, \dots, \mathbf{M}_N)$ 作为自编码器输入, 经过若干层的编码和解码操作, 得到由节点动态属性嵌入向量组成的网络动态属性嵌入矩阵 $\mathbf{H} = (h_1, h_2, \dots, h_n) \in \mathbb{R}^{n \times F}$ 作为模型的输出。

3.1.2 时序变化信息的提取

为了提取整个时序序列上机会网络单元间的时序变化信息, 采用类增量学习式的策略, 在训练的过程中充分利用历史的训练结果, 让时刻 τ 网络单元对应的嵌入模型继承时刻 $\tau-1$ 网络单元训练好的参数, 对第 t 个机会网络单元的网络信息进行迭代训练, 得到新的模型参数 θ_t 。进一步地, 在经过有限个该过程的迭代实现整个时序序列上时序信息的提取。采用该策略: 其一加快了整个动态网络嵌入模型的训练收敛速度, 使之更贴合机会网络的时变特性; 其二确保最终得到的节点动态嵌入矩阵 $\mathbf{H} = (h_1, h_2, \dots, h_n)$ 能够有效表征整个时序序列上网络的时序变化信息。

3.1.3 拓扑结构信息的提取

为了使节点动态属性嵌入向量更准确地表征网络拓扑结构信息, 构建相应的损失函数。通过不断最小化机会网络中节点的属性向量与重构向量之间的编码损失, 使得节点动态嵌入向量能够从局部结构以及全局结构上表征节点在网络中的拓扑信息。

在局部结构上, 自编码器的输入为节点属性向量 x_i , 它是网络属性矩阵中的第 i 行向量, 即该向量中的元素值代表着该节点与网络中其他节点的连接情况, x_i 中非零元的个数代表了节点在原始网络中的邻居数量。同理, 自编码器的输出(重构向量 \hat{x}_i)中的非零元个数代表了节点在重构网络中的邻居数量。而节点的邻居数量代表着节点的局部结构信息, 因此通过计算属性向量 x_i 与重构向量 \hat{x}_i 之间

的一阶范数(即反映节点的局部连接情况)构造局部结构损失函数 L_{loc} :

$$L_{\text{loc}} = \sum_{i=1}^N \|\hat{\mathbf{x}}_i - \mathbf{x}_i\|_1. \quad (9)$$

在全局结构上,节点属性向量 \mathbf{x}_i 与重构向量 $\hat{\mathbf{x}}_i$ 在网络空间上的欧氏距离反映了节点在网络全局结构上的差异^[23],因此通过计算属性向量 \mathbf{x}_i 与重构向量 $\hat{\mathbf{x}}_i$ 之间二阶范数(即向量之间的欧氏距离)构造全局损失函数:

$$L_{\text{glob}} = \sum_{i=1}^N \|\hat{\mathbf{x}}_i - \mathbf{x}_i\|_2^2, \quad (10)$$

加入 L_1 正则项,辅助模型产生稀疏矩阵,得到:

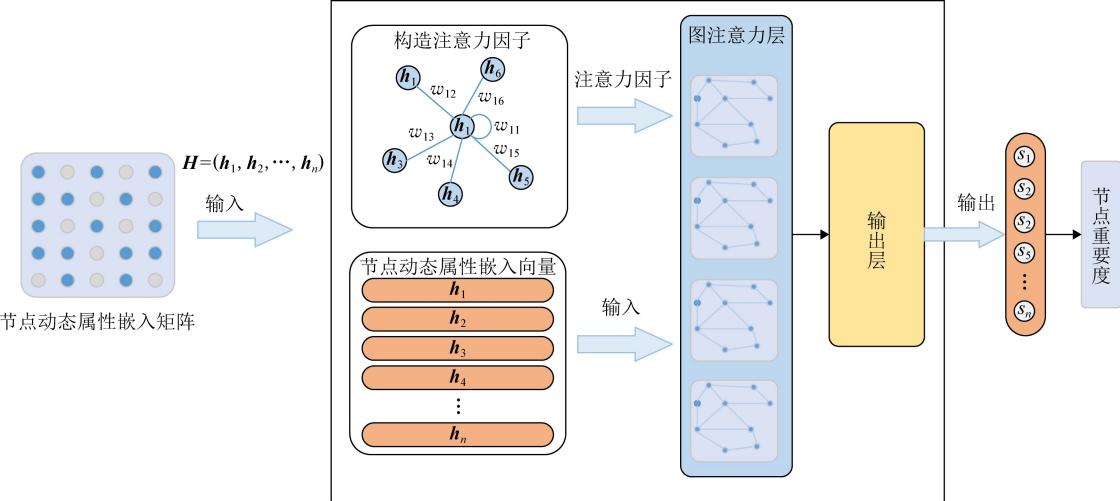


Fig. 8 Node importance evaluation model based on GNN

图 8 基于 GNN 的节点重要度评估模型

由节点动态属性嵌入向量组成的网络动态属性特征矩阵作为节点重要度评估模型的输入,节点重要度分数组成的重要度序列 $S = \{s_1, s_2, \dots, s_n\}$ 作为模型的输出。通过构造注意力因子,将节点动态属性嵌入矩阵 $H = (\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_n)$ 输入到图神经网络,节点与其邻居间的动态属性嵌入向量按照注意力因子进行聚合操作,从而对网络动态属性特征矩阵进行更新,得到图神经网络的输出,即融合了邻域信息的节点动态属性嵌入矩阵 $H' = (\mathbf{h}'_1, \mathbf{h}'_2, \dots, \mathbf{h}'_n) \in \mathbb{R}^{n \times F'}$ 。并将其作为输出层的输入,对更新后的节点动态嵌入向量进行降维转换,得到节点重要度分数组成的重要度序列 $S = \{s_1, s_2, \dots, s_n\}$ 。

3.2.1 图神经网络的确定

考虑到机会网络相邻节点间的重要度存在相互影响的情况,且节点间通信次数与通信时长会直接影响到节点间重要度的传递,为了有效地提取节点

$$L_1 = \sum_{l=1}^L (\|\mathbf{W}^{(l)}\|_1 + \|\hat{\mathbf{W}}^{(l)}\|_1), \quad (11)$$

得到模型的损失函数为

$$L = L_{\text{glob}} + \alpha L_{\text{loc}} + \beta L_1. \quad (12)$$

3.2 节点重要度评估模型的构建

利用 GNN 对图数据强大的特征自动化提取能力,避免传统方法人工构建网络中节点重要度特征。通过动态网络嵌入模型得到的节点动态属性嵌入向量组成的网络动态属性特征矩阵作为重要度评估模型的输入,以构建节点动态属性嵌入向量与节点重要度之间的映射关系。节点重要度评估模型如图 8 所示:

间重要度的相互影响关系,构建图神经网络将节点的动态属性嵌入向量按照节点间关系进行加权聚合,实现节点间邻域信息的融合。文献[24]提出了一种基于图数据的自注意力方法实现了一种根据节点间的关联关系对相邻节点进行卷积操作的图注意力网络(graph attention network, GAT)。引入该方法中所提出的 GAT 构建节点重要度评估模型。

3.2.2 注意力因子的构造

根据文献[24]所提出的图注意力网络,首先将节点动态属性嵌入矩阵 $H = (\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_n)$ 作为图注意力网络的输入,其中, $\mathbf{h}_i \in \mathbb{R}^F$, n 为节点总数, F 为节点动态属性嵌入向量的维数。使用 $\mathbf{h}'_i \in \mathbb{R}^{F'}$ 表示经过图注意力网络的输出, F' 为更新后的节点动态属性嵌入向量维数。对于节点动态属性嵌入向量 \mathbf{h}_i ,首先通过一个节点共享的特征转换矩阵 $\mathbf{W} \in \mathbb{R}^{F \times F'}$ 进行线性变换,再通过注意力机制 $\mathbf{a} \in \mathbb{R}^{2F'}$ 计算节

点间的注意力系数,此后通过 softmax 函数对节点与其邻居的注意力系数进行归一化操作得到节点间的注意力因子,最后利用注意力因子对节点的动态属性嵌入向量进行加权求和,得到融合了领域信息的新的节点动态属性嵌入矩阵 $\mathbf{H}' = (\mathbf{h}'_1, \mathbf{h}'_2, \dots, \mathbf{h}'_n)$.具体地,相邻节点对 (i, j) 之间的注意力系数为

$$e_{ij} = \mathbf{a}(\mathbf{W}\mathbf{h}_i, \mathbf{W}\mathbf{h}_j), \quad (13)$$

其中, \mathbf{a} 为单层前馈神经网络实现的注意力函数,通过 \mathbf{a} 计算相邻节点间的注意力系数 e_{ij} ,该系数表征了节点 j 与节点 i 的动态属性嵌入向量之间关联关系.

进一步地,为了方便计算节点 i 与其邻居节点之间的注意力系数,利用 softmax 函数对其进行归一化,得到相邻节点对 (i, j) 之间的注意力因子:

$$\alpha_{ij} = \text{softmax}_j(e_{ij}) = \frac{\exp(e_{ij})}{\sum_{k \in N_i} \exp(e_{ik})}, \quad (14)$$

其中, N_i 表示节点 i 的邻居节点集合,即机会网络中与节点 i 存在通信关系的节点集合.通过 α_{ij} 计算节点对 (i, j) 的动态属性嵌入向量的线性组合,从而对节点 i 的动态属性嵌入向量进行更新:

$$\mathbf{h}'_i = \sigma \left(\sum_{j \in N_i} \alpha_{ij} \mathbf{W}\mathbf{h}_i \right). \quad (15)$$

为了提高模型的性能,引入多头注意力机制:

$$\mathbf{h}'_i = \sigma \left(\frac{1}{K} \sum_{k=1}^K \sum_{j \in N_i} \alpha_{ij}^k \mathbf{W}^k \mathbf{h}_i \right), \quad (16)$$

其中, $\sigma(\cdot)$ 为激活函数, K 为多头注意力的头数, \mathbf{W}^k 是特征转换矩阵, α_{ij}^k 是计算第 k 个注意力的权重系数.通过对 K 个相互独立的注意力机制按照式(16)所示方式进行变换,得到最终的节点嵌入向量 \mathbf{h}'_i .

本文实验采用 LeakyReLU (leak 系数设置为 0.01) 作为注意力机制的非线性激活函数.根据文献 [24], 多头注意力机制的头数设置为 4.

通过图神经网络中注意力层的堆叠得到融合了领域信息的节点 i 的动态属性特征向量 \mathbf{h}'_i .最后,模型的输出层为 2 层的多层感知机,其输入为 $\mathbf{H}' = (\mathbf{h}'_1, \mathbf{h}'_2, \dots, \mathbf{h}'_n)$, 通过对 $\mathbf{H}' = (\mathbf{h}'_1, \mathbf{h}'_2, \dots, \mathbf{h}'_n)$ 进行降维转换, 得到对应的节点重要度分数序列 $S = \{s_1, s_2, \dots, s_n\}$.

3.2.3 损失函数

构建基于均方误差 (mean squared error, MSE) 的监督损失函数:

$$L = \frac{1}{N} \sum_{n=1}^N (s_n - g_n)^2, \quad (17)$$

其中, N 为样本总数, s_n 为节点 n 通过模型得到节

点重要度分数评估值, g_n 为节点 n 的重要度分数真实值.MSE 越小, 模型的评估值与真实值的差异越小, 对节点重要度评估越准确.

4 实验结果及分析

4.1 实验数据

本文实验在 CRAWDAD^[25] 提供的稀疏型机会网络、密集型机会网络以及大型机会网络的真实数据集上进行.3 个数据集的主要信息如表 1 所示:

Table 1 Trace Data Information

表 1 Trace 数据信息

数据集描述	MIT reality	Haggle 项目	Asturias-er
设备类型	诺基亚 6600	iMote	车载 GPS 设备
网络类型	蓝牙	蓝牙	GPS
持续时间/d	246	3	366
扫描间隔/s	300	120	30
有效(总)用户数	96	98	229
有效记录数	205187	341202	5245747

1) MIT reality 数据集.记录了 100 个携带 Nokia 6600 手机的用户自 2004-10—2005-05 共 246 d, 利用蓝牙接口相遇的数据(300 s 采集 1 次).

2) Haggle 项目数据集.收集了 Infocom2006 年会 98 位参会人员 3 d 携带蓝牙接口相遇的数据(120 s 采集 1 次).

3) Asturias-er 数据集.记录了 229 台车载无线通信设备或 GPS 设备在 2011-10—2012-10 期间生成的数据(30 s 采集 1 次).

本文采用 Pycharm 对机会网络数据集进行分析、处理,将其转化为网络属性矩阵.采用 Keras 框架作为动态网络嵌入模型以及节点重要度评估模型的训练工具.利用 Python 图形库 Matplotlib 对实验数据以及实验结果进行可视化.采用 Gephi 网络分析工具对机会网络数据集进行统计分析与可视化.

4.2 评价方法

采用 SI(susceptible infected), SIR(susceptible infected recovered) 传播模型评价和比较本文所提出的节点重要度评估方法;采用 NDCG@10 (normalized discounted cumulative gain) 对节点重要度排序列表的质量进行评价.

1) 基于传播模型的评价

通常认为,机会网络这类以消息传递为主的网络中,节点的重要度体现于节点在网络中对消息的

传播能力^[26].因此采用基于信息传播模型的评价方法对节点在网络中信息传递的重要度进行评价.利用 SI 与 SIR 感染模型^[26]中网络感染节点总数来衡量节点的重要度,SI 感染模型将网络中的节点分为 2 种状态,分别为易感染态和感染态,其中易感染态表示节点还未被感染疾病,但由于节点间的接触会被其他节点感染,感染态表示已经感染疾病的节点,易感染态节点在被感染节点感染后无法被恢复.SIR 模型则多出一种状态为免疫态,感染节点以恢复率恢复为免疫态.假设初始时刻 t_0 将总节点数为 n 的未感染网络中 i_0 个节点设置为感染源,剩余 $n - i_0$ 个节点为未感染节点,感染概率为 p .那么在时刻 i 感染节点总数为

$$i(t) = n \left[1 + \left(\frac{n}{i_0} - 1 \right) \exp(-pt) \right]. \quad (18)$$

具体地,在实验中,通过本文方法得到所有节点的重要度,并按照重要度大小对节点进行排序,得到排序列表中重要度排名靠前的节点作为传播模型的感染源,利用 SI 和 SIR 感染模型比较在同一时间内感染节点的数量.采用 SI 模型评价本文方法的合理性,采用 SIR 比较本文方法和其他方法.

2) 基于相关性的评价

归一化折扣累计增益(normalized discounted cumulative gain, NDCG)常用于评价排序列表的质量.该指标的取值范围为[0,1],值越大,说明给定的排序列表与理想排序列表越接近,即说明节点重要度评估方法越优.给定一个按照节点重要度分数的排序列表,排序列表中排名前 k 位的折扣累计增益为

$$DCG@k = \sum_{i=1}^k \frac{r_i}{\ln(i+1)}, \quad (19)$$

其中, r_i 为排序列表中位于第 i 位节点的重要度分数,考虑到理想排序列表(即按照节点真实重要度分数得到的排序列表)中前 k 位的理想折扣累计增益($IDCG@k$),进行归一化操作得到归一化折扣累计增益为

$$NDCG@k = \frac{DCG@k}{IDCG@k}. \quad (20)$$

4.3 实验参数设置

实验参数设置包括机会网络单元切片时长 Δt 的设置和动态网络嵌入模型中超参数的设置.

1) 机会网络单元切片时长 Δt

考虑机会网络数据集持续性与周期性^[27],确定

MIT, Haggle, Asturias-er 的候选切片时长分别为 10 d, 4 h, 48 h.以候选切片时长为基准,按照 $\Delta t \in \{8 \text{ d}, 9 \text{ d}, 10 \text{ d}, 11 \text{ d}, 12 \text{ d}\}$, 在 MIT 数据集上,根据 Δt 对排序质量 $NDCG@10$ 的影响,最终确定切片时长,实验结果如图 9 所示.其中 $\Delta t = 11 \text{ d}$ 时 $NDCG@10$ 取值最优.因此,本文选取机会网络单元切片时长 $\Delta t = 11 \text{ d}$ 作为 MIT 的切片时长.

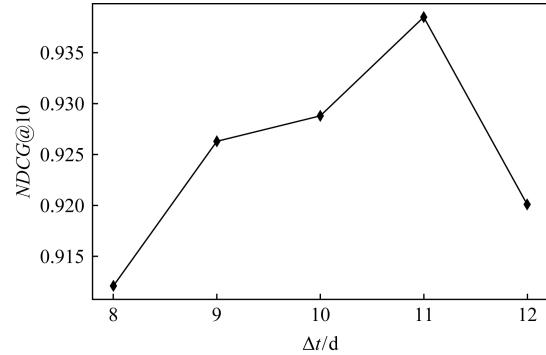


Fig. 9 Influence of Δt on $NDCG@10$ of MIT
图 9 MIT 数据集上 Δt 对 $NDCG@10$ 的影响

以候选切片时长 $\Delta t = 2 \text{ h}$ 为基准,按照 $\Delta t \in \{2 \text{ h}, 3 \text{ h}, 4 \text{ h}, 5 \text{ h}, 6 \text{ h}\}$,在 Haggle 数据集上,根据 Δt 对排序质量 $NDCG@10$ 的影响,最终确定切片时长.实验结果如图 10 所示.其中 $\Delta t = 3 \text{ h}$ 时 $NDCG@10$ 取值最优.因此,本文选取 $\Delta t = 3 \text{ h}$ 作为 Haggle 的切片时长.

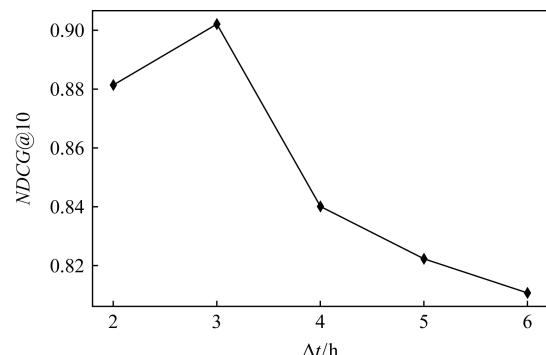
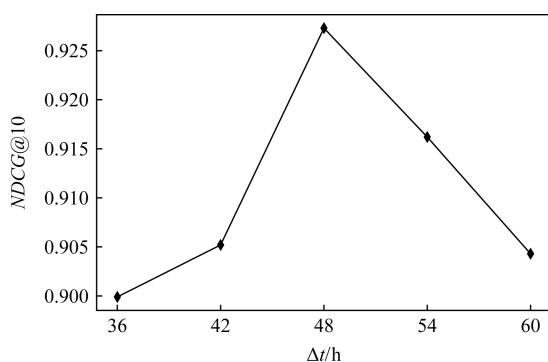


Fig. 10 Influence of Δt on $NDCG@10$ of Haggle
图 10 Haggle 数据集上 Δt 对 $NDCG@10$ 的影响

以候选切片时长 $\Delta t = 48 \text{ h}$ 为基准,按照 $\Delta t \in \{36 \text{ h}, 42 \text{ h}, 48 \text{ h}, 54 \text{ h}, 60 \text{ h}\}$,在 Asturias-er 数据集上,根据 Δt 对排序质量 $NDCG@10$ 的影响,最终确定切片时长.实验结果如图 11 所示.其中 $\Delta t = 48 \text{ h}$ 时 $NDCG@10$ 取值最优.因此,本文选取 $\Delta t = 48 \text{ h}$ 为 Asturias-er 的切片时长.

Fig. 11 Influence of Δt on NDCG@10 of Asturias-er图 11 Asturias-er 数据集上 Δt 对 NDCG@10 的影响

2) 动态网络嵌入模型的超参数

在基于自编码器的动态网络嵌入模型中,超参数的具体配置为:网络嵌入的最佳网络嵌入维数与网络的规模和密度相关^[27],考虑到所选机会网络数据集的规模(节点数目均小于300)以及机会网络的稀疏性(网络密度较小),将MIT, Haggle, Asturias-er的网络嵌入维数设置为 $F=32$;考虑到网络中节点的全局结构相较于局部结构更能反映节点在网络中的重要程度^[28]且正则项主要是辅助模型的训练,故设置式(12)中的 $\alpha=10^{-5}$, $\beta=10^{-4}$.采用Adam学习算法对自编码器进行训练,迭代次数设置为200,编码器与解码器隐藏层数量均为4,每层的神经元数量分别为 $n, 84, 64, 32$ (n 为网络节点数量),学习率参数设置为0.001.

4.4 实验结果

实验分为2个部分:1)通过信息传播模型评价节点在机会网络中的消息传播能力;2)通过相关性分析,利用NDCG@10衡量本文方法与其他方法所得到的节点重要度排序列表的质量.

1) 节点在机会网络中的消息传播能力

机会网络中节点的核心任务是消息的投递,网络中不同个体的社会属性决定了其在网络中对于消息的传播能力,而在网络中节点对消息的传播能力可分为对消息的传播速率和覆盖范围.本文提出的机会网络节点重要度评估方法,目的是找到并利用消息传播能力较高的节点以提高整个网络的消息传播效率.因此,本文通过实验1验证节点重要度高的节点对于消息的传播速率影响,通过实验2验证节点重要度高的节点对于消息覆盖范围的影响.

实验1.采用SI传播模型考察本文方法得到的节点重要度列表中排名不同的节点的消息传播速率,进而说明本文方法的有效性.为了直观地展示网

络中不同节点的消息传播速率,选取节点重要度排序列表中间隔相同的5个节点分别单独作为消息源对比其消息传播速率.若排名靠前的节点在消息的传播速率(即图12中曲线下的覆盖面积)上要高于排名靠后的节点,则说明重要度高的节点能够更快地使消息传播到更多的节点上,进而验证本文方法

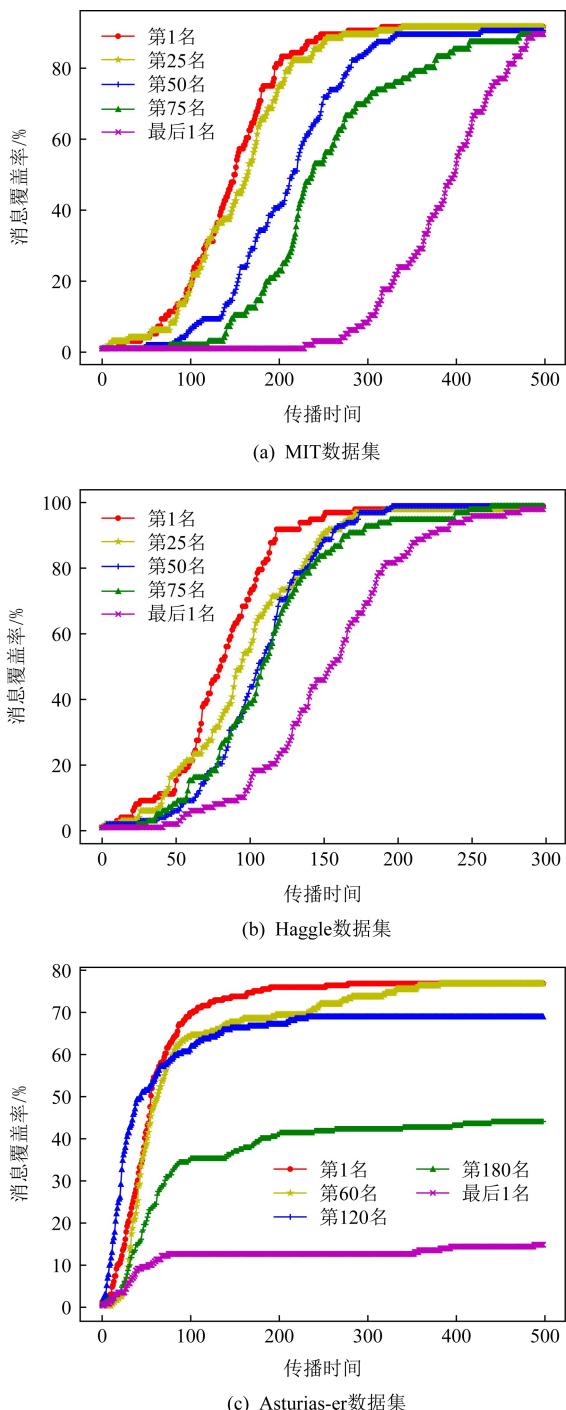


Fig. 12 Comparison of network message propagation rate

图 12 网络消息传播速率对比

对节点重要度的评估是有效的。具体地,根据经验值设定 SI 传播模型的传播概率,观察消息的传播速率。图 12(a)为 MIT 数据集下,选取节点重要度排序列表中第 1(即 top 节点)、第 25、第 50、第 75 以及最后 1 名(即 bottom 节点)的 5 个节点分别作为 SI 传播模型的传播源的消息传播速率情况。从实验结果可知,由于 MIT 数据集持续时间较长,且有效记录数较少,导致网络连接较为稀疏,将单一节点作为消息源时,消息能够覆盖全网 85% 左右,top 节点的消息传播速度明显快于排名靠后的节点。从整体趋势来看,排名靠前的节点其消息传播速率上要快于排名靠后的节点,从逐个节点传播情况来看,top 节点的消息传播速率明显快于 bottom 节点,且节点重要度排序越靠后其消息覆盖速率越低。图 12(b)为 Haggle 数据集下,同样选取节点重要度排序列表中第 1(top)、第 25、第 50、第 75 以及最后 1 名(bottom)的 5 个节点进行实验比较不同节点的消息传播速率。从实验结果可知,相比于 MIT 而言,Haggle 持续时间短且有效记录数多,节点间连接较为密集,因此,消息能够迅速覆盖到整个网络。总体上来说,top 节点的消息传播速率均快于其他节点,且排名越靠后其消息传播速率越慢,但由于 Haggle 数据集较 MIT 数据集更密集,使得网络中节点的传播速率总体上要更高,且 top 节点与 bottom 节点的差距较 MIT 也更小。图 12(c)为 Asturias-er 数据集下,选取节点重要度排序列表中第 1(top)、第 60、第 120、第 180 以及最后 1 名(bottom)节点作为实验对象考察其消息传播速率。从实验结果可知,由于 Asturias-er 节点数目多且持续时间长,不同节点间的消息传播能力存在明显差距。从整体上来说,排名越靠后的节点其在消息传播速率以及消息覆盖率上都明显低于排名靠前的节点,且排名越靠后其差距越大,从节点层面上来说,top 节点与 bottom 节点间无论是在消息覆盖范围和消息传播速率上都有明显的差距,top 节点的消息覆盖率能达到 80% 左右而 bottom 节点的消息覆盖率不足 20%,且在传播过程中 bottom 节点消息传播速率也要显著慢于 top 节点。

综上所述,在 MIT, Haggle, Asturias-er 数据集上,实验结果表明按照本文方法得到的节点重要度由高向低将节点进行排序,等差选取 5 个节点单独作为 SI 模型的传播源时,排名靠前的节点消息传播速率高于排名靠后的节点。因此,本文方法得到的节点重要度能够准确地描述节点对于消息的传播速率,也说明了本文方法是有效的。

实验 2.采用 SIR 模型比较本文方法和时效度(temporal degree, TD)方法^[11]、时效介数(temporal betweenness, TB)方法^[12]、时效 PageRank (temporal PageRank, f-PageRank)方法^[29]以及 kshell-CN 方法^[30],验证模型的正确性。具体地,选取时 TD, TB, f-PageRank, kshell-CN 以及本文方法(以下简称 GNN)得到的节点重要度排序列表中排名前 5 的节点作为重要节点集合 S, 分别将不同方法得到的 top1, top2, top3, top4, top5, S 作为 SIR 传播模型的传播源,比较不同方法的消息覆盖范围。相同的传播时间步下,消息覆盖节点数越多,则说明节点重要度评估方法越准确。

表 2~4 分别给出了在 MIT, Haggle, Asturias-er 机会网络数据集下由 TD, TB, f-PageRank, kshell-CN, GNN 节点重要度评估方法得到的排名 top1~5 节点以及由这些节点组成的重要节点集合 S。

**Table 2 MIT Node Importance Rank Table
表 2 MIT 节点重要度排序表**

方法	top1	top2	top3	top4	top5	S
TD	28	56	42	89	85	{28,56,42,89,85}
TB	28	42	56	95	85	{28,42,56,95,85}
f-PageRank	56	42	95	85	89	{56,42,95,85,89}
kshell-CN	28	56	95	85	89	{28,56,95,85,89}
GNN	89	56	80	85	95	{89,56,80,85,95}

**Table 3 Haggle Node Importance Rank Table
表 3 Haggle 节点重要度排序表**

方法	top1	top2	top3	top4	top5	S
TD	86	46	44	88	77	{86,46,44,88,77}
TB	82	15	96	77	46	{85,15,96,77,46}
f-PageRank	82	86	77	88	44	{82,86,77,88,44}
kshell-CN	44	46	77	86	88	{44,46,77,86,88}
GNN	77	72	44	92	85	{77,72,44,92,85}

**Table 4 Asturias-er Node Importance Rank Table
表 4 Asturias-er 节点重要度排序表**

方法	top1	top2	top3	top4	top5	S
TD	33	158	109	34	44	{33,158,109,34,44}
TB	33	158	44	104	35	{33,158,44,104,35}
f-PageRank	33	158	104	44	34	{33,158,104,44,34}
kshell-CN	158	33	109	34	113	{158,33,109,34,113}
GNN	33	158	104	27	35	{33,158,104,27,35}

首先,以 S 为传播源考察 3 个数据集下消息覆盖

范围结果,图 13 为实验结果。从图 13(a)中可以看出,由于 MIT 数据集属于稀疏型机会网络,节点间连接频率低,整体的消息覆盖节点数较少,本文方法得到的重要节点集合 S 作为消息源时,网络的消息覆盖节点数要明显高于 TD, TB 方法,略高于 f-PageRank 方法与 kshell-CN 方法。图 13(b)为 Haggle 数据集网络消息覆盖范围结果,由于 Haggle 数据集属于

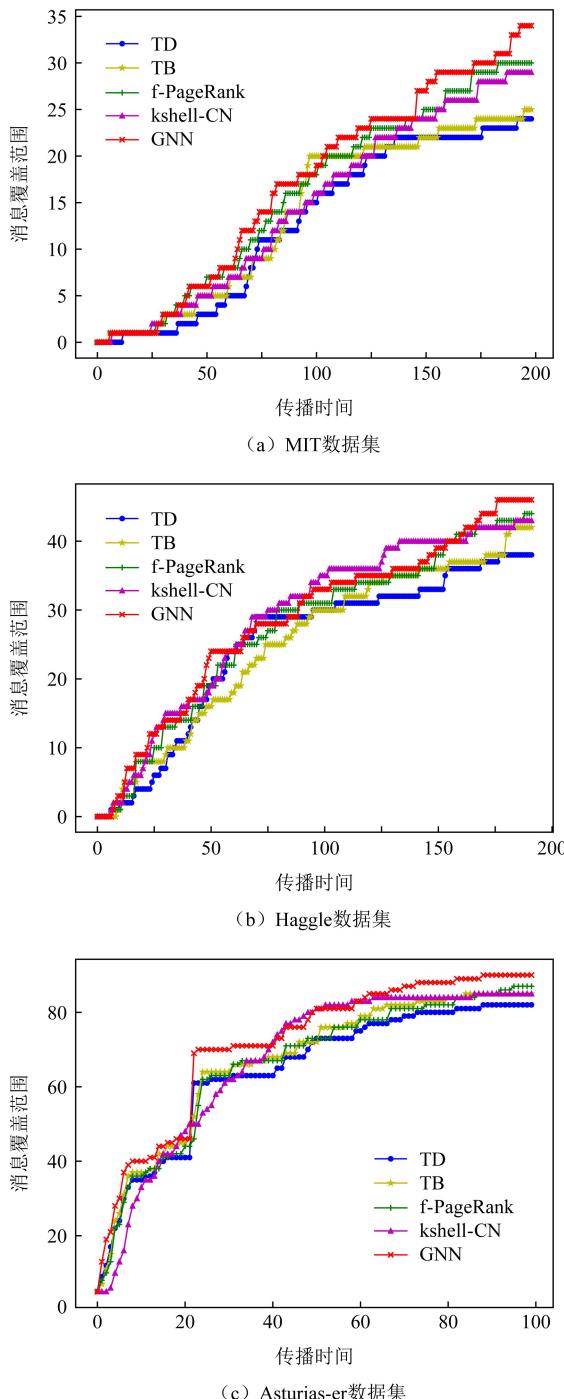


Fig. 13 Comparison of network message coverage

图 13 网络消息覆盖范围对比

密集型机会网络,节点连接频繁,即便在相同的传播条件下,4 种方法所得到的消息覆盖范围更广。由本文方法得到的 S 作为消息源时,网络的消息覆盖率明显高于 TD 方法,略高于 TB,kshell-CN,f-PageRank 方法。图 13(c)为 Asturias-er 数据集网络消息覆盖范围,Asturias-er 数据集持续时间长,且随着时间的推移参与通信的节点数目也随之增加。总体来说,以本文方法得到的重要节点集合 S 作为消息源时网络的消息覆盖范围均高于其他方法。

其次,逐一考察由不同方法得到的 top1~5 节点在 SIR 传播模型下的消息覆盖范围。图 14~16 分别为 MIT,Haggle,Asturias-er 数据集下,不同方法 top1~5 节点在 SIR 传播模型下的消息覆盖范围对比图。

图 14 为 MIT 数据集下 top1~5 的对比结果。其中,图 14(a)是不同方法 top1 节点的对比情况,根据表 2 结果选取 89,56,28 号节点单独作为 SIR 传播模型的传播源进行实验,从图 14(a)中可知,89 号节点优于 56 号节点与 28 号节点。类似地,图 14(b)~(e)为按照表 2 结果分别比较 top2~5 不同方法得到的节点的消息覆盖范围。特别地,top4 和 top5 节点集中在 85 和 95 号节点上,故图 14(d)和图 14(e)均为 85 号和 95 号节点的对比情况。从实验结果可知,本文方法在 top1,top3,top4,top5 的评估更为准确。

图 15 为 Haggle 数据集下 top1~5 的对比结果。其中,图 15(a)是不同方法 top1 节点的对比情况,根据表 3 的结果选取 77,86,44 号节点单独作为 SIR 传播模型的传播源进行实验,从图 15(a)中可知,77 号节优于 86 与 44 号节点。类似地,图 15(b)~(e)分别为 top2~5 不同方法得到的节点消息覆盖范围。从实验结果可知,本文方法在 top1~5 的评估更为准确。

图 16 为 Asturias-er 数据集下 top1~5 的对比结果。其中,图 16(a)和图 16(b)是不同方法 top1 与 top2 节点的对比情况,根据表 4 不同方法的评估结果中 top1 和 top2 的节点均是 33 号和 158 号节点,故选取 33,158 号节点单独作为 SIR 传播模型的传播源进行实验,从图 16(a)中可知,33 号节点与 158 号节点消息覆盖率基本一致,因此多种评估方法的 top1 与 top2 节点均是 158 号和 33 号节点。类似地,图 15(c)(d)(e)分别为 top3~5 不同方法得到的节点消息覆盖范围。从实验结果可知,本文方法在 top1,top2,top3,top4 的评估更为准确。

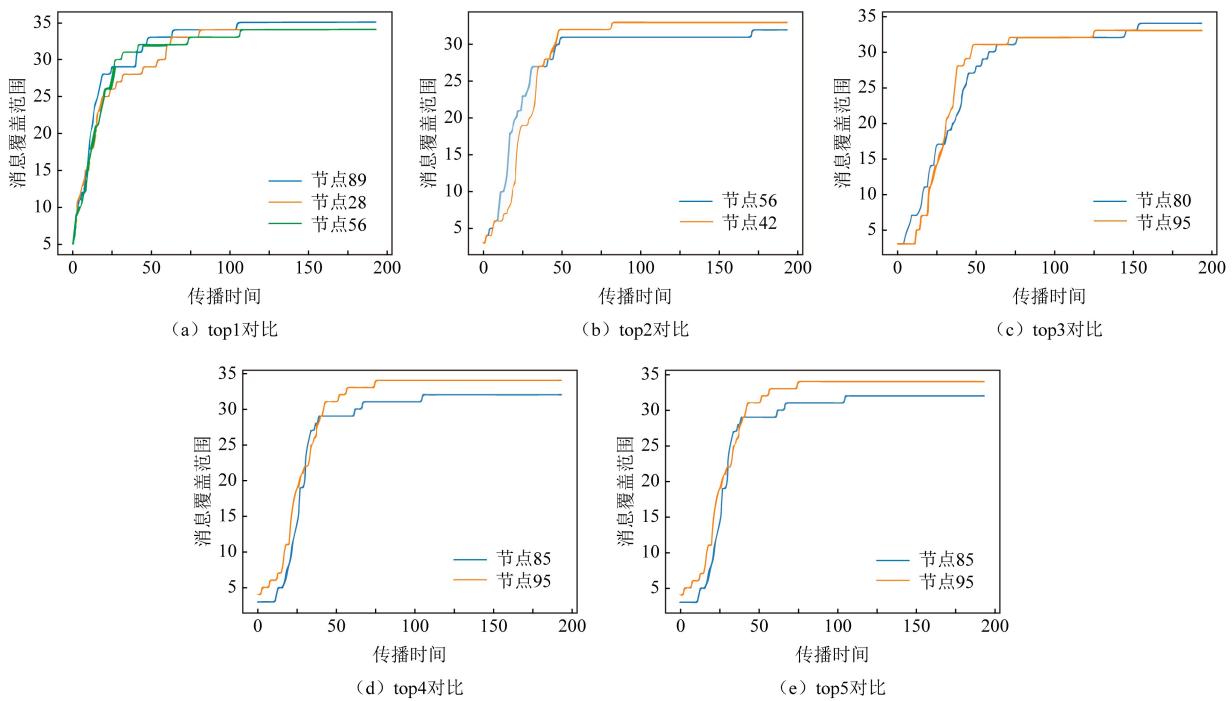


Fig. 14 Comparison of different methods top1~5 under MIT

图 14 MIT 下不同方法 top1~5 对比

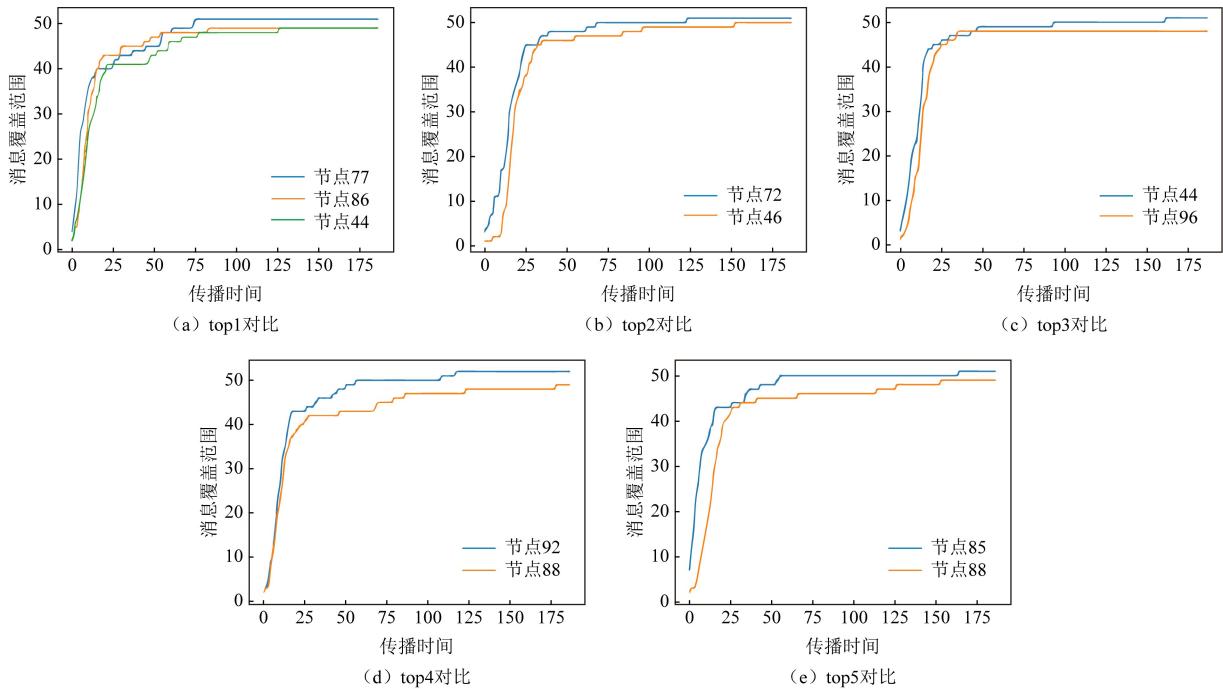


Fig. 15 Comparison of different methods top1~5 under Haggle

图 15 Haggle 下不同方法 top1~5 对比

综上所述,在 MIT, Haggle, Asturias-er 数据集上,本文方法得到的 top1~5 节点以及由它们组成的重要节点集合 S,在 SIR 传播模型上的消息覆盖率均高于 TD 方法、TB 方法、f-PageRank 方法以及

kshell-CN 方法。这是由于本文方法综合考虑了节点的局部拓扑信息、全局拓扑信息、时序变化信息以及节点间重要度的相互影响关系,因此,能更加准确地识别出机会网络中的重要节点。

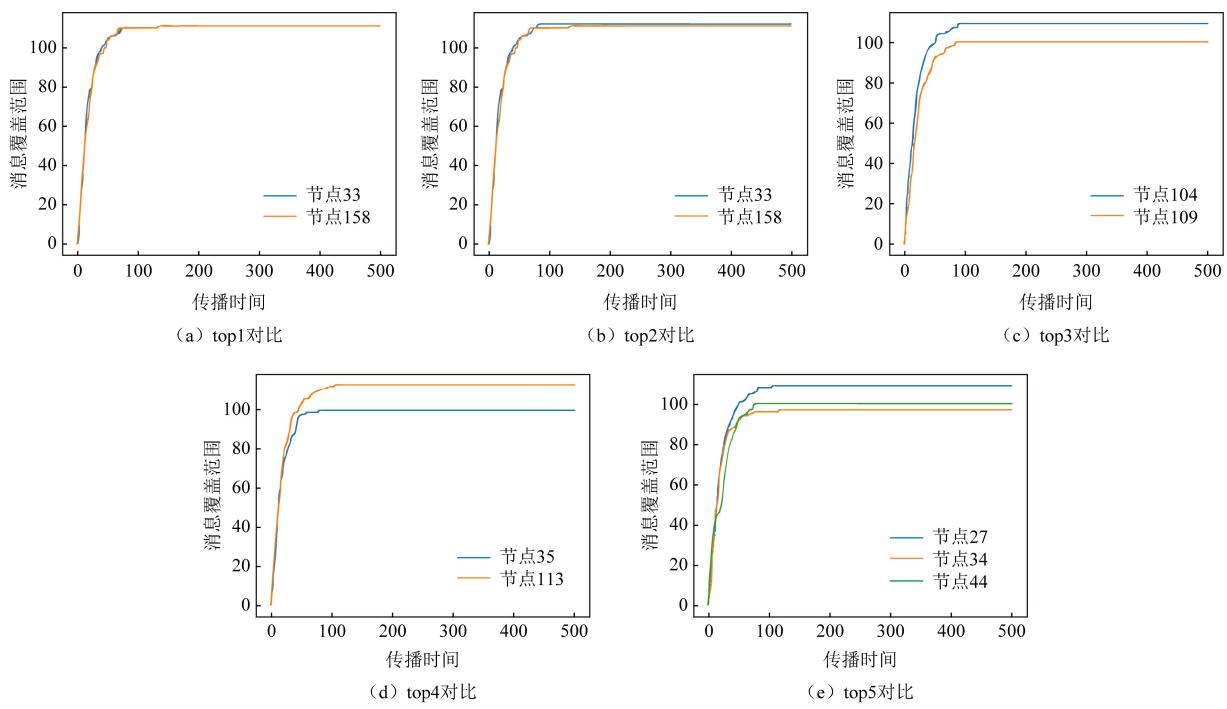


Fig. 16 Comparison of different methods top1~5 under Asturias-er

图 16 Asturias-er 下不同方法 top1~5 对比

2) 节点重要度排序列表的质量

根据 TD 方法、TB 方法、f-PageRank 方法、kshell-CN 方法及本文方法得到节点重要度的排序列表, 将消息传播模型的结果作为标准排序。分别计算不同网络数据集中 TD 方法、TB 方法、f-PageRank 方法、kshell-CN 方法以及本文方法的 NDCG@10, 排序质量如表 5 所示:

Table 5 NDCG@10 of Three Datasets

表 5 3 个数据集上的 NDCG@10 值

方法	MIT	Haggle	Asturias-er
TD	0.7676	0.7968	0.8052
TB	0.7688	0.8737	0.8857
f-PageRank	0.8763	0.8991	0.9108
kshell-CN	0.8543	0.8883	0.8889
GNN	0.9385	0.9021	0.9273

注:黑体数字表示实验结果中的最优值。

由表 5 可知,由于 MIT 数据集持续时间长,且有效记录数较少,属于稀疏性机会网络,采用 TD 与 TB 方法对节点重要度评估时,导致网络的拓扑结构信息以及时序变化信息不能被充分利用,本文方法在排序质量上相较于 TD 方法与 TB 方法分别提高了 22.2% 与 22.0%,相较于 f-PageRank 方法提高了 7.1%。在 Haggle 数据集中,由于 Haggle 持续时

间短,节点间通信频繁,属于密集型机会网络。随着节点间的有效通信路径增加,TB 方法能够更好地刻画出节点对于网络中有效通信路径的控制力,因此,评估结果较 TD 更好。但相较本文方法,TB 方法没有考虑节点间重要度的相互影响,故评估效果低于本文方法。在 Asturias-er 数据集中,由于 Asturias-er 节点规模相较于其他 2 个数据集更大且数据集持续时间长,因此能够获取的网络拓扑信息以及时序变化信息更加丰富,本文方法相较于其他方法能够更有效地利用这些信息进行节点重要度的评估,故评估效果均高于其他方法。基于以上分析,本文所提出的方法能够有效地对机会网络中的节点重要度进行评估,且实验结果表明本文方法在多个评价方法上优于现有其他方法。

5 总 结

本文方法针对机会网络的动态性与时变性,利用机会网络数据的历史连接信息有效地对机会网络中节点的重要度进行评估,从而识别出机会网络中消息传播能力强的节点。实验结果表明:本文方法能够准确识别出在机会网络中消息传播能力强的节点,将这些节点作为消息的传播源时,能提高消息覆盖率和传播速率。同时,在稀疏型机会网络场景、密集型

机会网络场景以及大型机会网络场景下,本文提出的基于GNN的机会网络节点重要度评估方法,在SI,SIR和 $NDCG@10$ 评价方法上优于TD,TB,f-PageRank,kshell-CN。未来研究将进一步确定机会网络单元的最佳切片时长以尝试改善评估的效果。

作者贡献声明:刘琳岚负责理论模型构建、实验方案设计和文章撰写;谭镇阳提出研究思路,负责实验数据整理与分析、实验结果可视化,撰写初稿;舒坚参与了论文想法的讨论,确定研究思路,对于实验方案提出指导意见,审阅和完善论文内容。

参 考 文 献

- [1] Xiong Yongping, Sun Limin. Opportunistic network [J]. Journal of Software, 2009, 20(1): 124–137 (in Chinese)
(熊永平, 孙利民. 机会网络[J]. 软件学报, 2009, 20(1): 124–137)
- [2] Wu Yue, Li Jianhua, Lin Chuang. Survey of security and trust in opportunistic networks [J]. Journal of Computer Research and Development, 2013, 50(2): 278–290 (in Chinese)
(吴越, 李建华, 林闯. 机会网络中的安全与信任技术研究进展[J]. 计算机研究与发展, 2013, 50(2): 278–290)
- [3] Li Jie, Hong Tao, Wang Xingwei, et al. A data dissemination mechanism based on group structure in opportunistic mobile social networks [J]. Journal of Computer Research and Development, 2019, 56(11): 2494–2505 (in Chinese)
(李婕, 洪韬, 王兴伟, 等. 机会移动社交网络中基于群组构造的数据分发机制[J]. 计算机研究与发展, 2019, 56(11): 2494–2505)
- [4] Ren Xiaolong, Lü Linyuan. A survey of the scheduling methods of important nodes in network [J]. Chinese Science Bulletin, 2014, 59(13): 1175–1197 (in Chinese)
(任晓龙, 吕琳媛. 网络重要节点排序方法综述[J]. 科学通报, 2014, 59(13): 1175–1197)
- [5] Alsayed A, Higham D J. Betweenness in time dependent networks [J]. Chaos, Solitons & Fractals, 2015, 72(4): 35–48
- [6] Tang J, Musolesi M, Mascolo C, et al. Analysing information flows and key mediators through temporal centrality metrics [C] //Proc of the 3rd Workshop on Social Network Systems. New York: ACM, 2010: 1–6
- [7] Kitsak M, Gallos L K, Havlin S, et al. Identification of influential spreaders in complex networks [J]. Nature Physics, 2010, 6(11): 888–893
- [8] Zareie A, Sheikhahmadi A. A hierarchical approach for influential node ranking in complex social networks [J]. Expert Systems with Applications, 2017, 93(4): 200–211
- [9] Wang Zhixiao, Zhao Ya, Xi Jingke, et al. Fast ranking influential nodes in complex networks using a k -shell iteration factor [J]. Physica A: Statistical Mechanics and Its Applications, 2016, 29(5): 171–181
- [10] Zareie A, Sheikhahmadi A, Jalili M. Influential node ranking in social networks based on neighborhood diversity [J]. Future Generation Computer Systems, 2019, 94(5): 120–129
- [11] Chen Shi, Ren Zhuoming, Liu Chuang, et al. Identification methods of vital nodes on temporal networks [J]. Journal of University of Electronic Science and Technology of China, 2020, 49(2): 291–314 (in Chinese)
(陈诗, 任卓明, 刘闯, 等. 时序网络中关键节点的识别方法研究进展[J]. 电子科技大学学报, 2020, 49(2): 291–314)
- [12] Eppstein D, Goodrich M T, Strash D, et al. Extended h-Index parameterized data structures for computing dynamic subgraph statistics [J]. Theoretical Computer Science, 2010, 447(12): 128–141
- [13] Wen Xiangxi, Tu Conglinag, Wu Minggong, et al. Fast ranking nodes importance in complex networks based on LS-SVM method [J]. Physica A: Statistical Mechanics and Its Applications, 2018, 506: 11–23
- [14] Wang Dingjie, Zou Xiufen. A new centrality measure of nodes in multilayer networks under the framework of tensor computation [J]. Applied Mathematical Modelling, 2018, 54(7): 46–63
- [15] Wang Huifang, Zhang Chenyu, Lin Dongyang, et al. An artificial intelligence method based on network embedding and support vector regression for evaluating the importance of power grid nodes [J]. Frontiers of Information Technology & Electronic Engineering, 2019, 20(6): 816–828 (in Chinese)
(王慧芳, 张晨宇, 林东阳, 等. 用于电网节点重要度评估的一种基于网络嵌入和支持向量回归的人工智能方法[J]. 信息与电子工程前沿, 2019, 20(6): 816–828)
- [16] Park N, Kan A, Dong Xinluna, et al. Estimating node importance in knowledge graphs using graph neural networks [J]. arXiv preprint, arXiv: 1905.08865, 2019
- [17] Sun Qindong, Qiao Yimin, Wang Jiamin, et al. Node importance evaluation method in wireless sensor network based on energy field model [J]. EURASIP Journal on Wireless Communications and Networking, 2016, 2016(1): 199–206
- [18] Han Zhongming, Mao Rui, Zheng Chenye, et al. An effective influence model of dynamic network nodes [J]. Application Research of Computers, 2019, 36(7): 1960–1964 (in Chinese)
(韩忠明, 毛锐, 郑晨烨, 等. 一种有效的动态网络节点影响力模型[J]. 计算机应用研究, 2019, 36(7): 1960–1964)
- [19] Qiu Jiezhong, Tang Jian, Ma Hao, et al. Deepinf: Social influence prediction with deep learning [C] //Proc of the 24th ACM SIGKDD Int Conf on Knowledge Discovery & Data Mining. New York: ACM, 2018: 2110–2119

- [20] Gou Chengcheng, Du Pan, He Min, et al. Tsk-shell: An algorithm for finding topic-sensitive influential spreaders [J]. Journal of Computer Research and Development, 2017, 54(2): 361–368 (in Chinese)
(苟程成, 杜攀, 贺敏, 等. Tsk-shell: 一种话题敏感的高影响力传播者发现算法[J]. 计算机研究与发展, 2017, 54(2): 361–368)
- [21] Zheng Wenping, Wu Zhikang, Yang Gui. A novel algorithm for identifying critical nodes in networks based on local centrality [J]. Journal of Computer Research and Development, 2019, 56(9): 1872–1880 (in Chinese)
(郑文萍, 吴志康, 杨贵. 一种基于局部中心性的网络关键节点识别算法[J]. 计算机研究与发展, 2019, 56(9): 1872–1880)
- [22] Clauset A, Eagle N. Persistence and periodicity in a dynamic proximity network [J]. arXiv preprint, arXiv: 1211.7343, 2012
- [23] Goyal P, Kamra N, He Xinren, et al. DynGEM: Deep embedding method for dynamic graphs [J]. arXiv preprint, arXiv: 1805.11273, 2018
- [24] Velickovic P, Cucurull G, Casaniva A, et al. Graph attention network [J]. arXiv preprint, arXiv: 1710.10903, 2019
- [25] CRAWDAD: A community resource for archiving wireless data [EB/OL]. Hanover, New Hampshire: Dartmouth College Libraries, 2014 [2020-01-10]. <http://www.crawdad.org>
- [26] López M, Peinado A, Ortiz A. An extensive validation of a SIR epidemic model to study the propagation of jamming attacks against IoT wireless networks [J]. Computer Networks, 2019, 165: 106945
- [27] Gu Weiwei, Tandon A, Ahn Y Y, et al. Defining and identifying the optimal embedding dimension of networks [J]. arXiv preprint, arXiv: 2004.09928, 2020
- [28] Tu Cunchao, Zhang Zhengyan, Liu Zhiyuan, et al. TransNet: Translation-based network representation learning for social

relation extraction [C] //Proc of the 26th Int Joint Conf on Artificial Intelligence. Melbourne, Australia: IJCAI, 2017: 2864–2870

- [29] Lü Laishui, Zhang Kun, Zhang Ting, et al. PageRank centrality for temporal networks [J]. Physics Letters A, 2019, 383(12): 1215–1222
- [30] Li Chao, Wang Li, Sun Shiwen, et al. Identification of influential spreaders based on classified neighbors in real-world complex networks [J]. Applied Mathematics and Computation, 2018, 320(3): 512–523



Liu Linlan, born in 1968. BSc. Professor. Member of CCF. Her main research interests include software engineering and distributed system.

刘琳岚,1968 年生.学士,教授,CCF 会员.主要研究方向为软件工程和分布式系统.



Tan Zhenyang, born in 1995. Master candidate. Student member of CCF. His main research interest is opportunistic networks.

谭镇阳,1995 年生.硕士研究生.CCF 学生会员.主要研究方向为机会网络.



Shu Jian, born in 1964. MSc. Professor. Senior member of CCF. His main research interests include wireless sensor networks, embedded system and software engineering.

舒 坚,1964 年生.硕士,教授,CCF 高级会员.主要研究方向为无线传感器网络、嵌入式系统和软件工程.