

## 在线社会关系语义化度量的可视化分析

王 渊<sup>1,2</sup>, 孙樱郡<sup>1,2</sup>, 王 博<sup>1,2</sup>, 杨朝洲<sup>1</sup>, 杨 量<sup>1</sup>

(1.天津大学计算机科学与技术学院, 天津 300000; 2.天津大学天津市认知计算与应用重点实验室, 天津 300000)

**摘要:** 传统的社会关系分析中, 社会关系的属性被视为客观且独立于参与者主观认知的。然而, 在涉及人物主观行为分析的社会计算问题中, 关系参与者对关系属性的主观认知时常比关系属性的客观值更加重要。为此尝试利用人们之间的互动语言内容, 来可视化刻画和分析关系双方对关系强度的主观认知。从互动语言中提取 4 种可以用于刻画互动语言关键的语言特征, 包括频度、长度、规范度和情感极性, 并通过将个人语言习惯因素从互动语言特征值中排除, 更加客观地反映关系参与者对关系的主观认知。在量化基础上, 从不同的观察角度, 设计并实现了一套较为完整的社会关系语义化度量可视化展示及分析系统。最后, 以电子邮件数据为例, 验证并展示了社会关系语义化度量的可视化结果, 并基于可视化结果初步分析了影响社会关系语义化属性的若干因素。

**关键词:** 社会关系分析; 社会网络; 互动语言; 可视化分析; 语义化度量

**中图分类号:** TP391 **文献标识码:** A **doi:** 10.3969/j.issn.0253-2778.2018.04.007

**引用格式:** 王渊, 孙樱郡, 王博, 等. 在线社会关系语义化度量的可视化分析[J]. 中国科学技术大学学报, 2018, 48(4):314-321.

WANG Yuan, SUN Yingjun, WANG Bo, et al. Visual analysis of semantic measuring of online social relationships[J]. Journal of University of Science and Technology of China, 2018, 48(4):314-321.

## Visual analysis of semantic measuring of online social relationships

WANG Yuan<sup>1,2</sup>, SUN Yingjun<sup>1,2</sup>, WANG Bo<sup>1,2</sup>, YANG Chaozhou<sup>1</sup>, YANG Liang<sup>1</sup>

(1. School of Computer Science and Technology, Tianjin University, Tianjin 300000, China;

2. Tianjin Key Laboratory of Cognitive Computing and Application, Tianjin 300000, China)

**Abstract:** In the traditional social relationships analysis, the attribute of social relation is regarded as objective and independent of the subjective cognition of the participant. However, in the social computing studies related to subjective behaviors, subjective features are often more important than objective features. The semantics of social relationship with the interactive language between individuals is visualized. Based on The key features of the interactive language in the theory of social linguistics, four language features to describe the semantics of social relationships are calculated, including frequency, length, fluency and sentiment polarity. By measuring and distinguishing personal language habits, the semantic measuring of social relationship more appropriate. To make semantic measuring more understand, a visual analysis system is implemented for online social relationships using Email data as a

**收稿日期:** 2017-05-31; **修回日期:** 2017-06-25

**基金项目:** 国家自然科学基金联合培育项目(U1736219), 国家自然科学基金联合重点项目(U1636203), 国家重点研发计划(2017YFE0111900), 天津市自然科学基金青年项目(14JQCQJC00400), 国家社科重大项目(14ZDB153), 国家重点基础研究发展计划(973)(2013CB329304)资助。

**作者简介:** 王渊, 女, 1994年生, 硕士研究生, 研究方向: 社会计算. E-mail: 939931891@qq.com

**通讯作者:** 王博, 博士/副教授. E-mail: bo\_wang@tju.edu.cn

case, and the factors related to the semantic features of online social relationships are primarily analyzed.

**Key words:** social relationships analysis; social network; interactive language; visualization; semantic measuring

## 0 引言

社会关系是构成人类社会的基本因素,互联网环境下的在线社会是真实社会的平行系统<sup>[1]</sup>.近年来,在线社会关系的建模与分析研究取得了长足进展,涉及的主要信息主要包括关系参与者的人物属性、关系拓扑特征以及人物间的互动信息.虽然对上述信息的应用方式不同,但当前的研究多存在关系客观性假设:假定人际关系的属性(例如方向、强度和类型)是客观的,独立于关系双方的观察角度.关系属性可以在不考察关系双方主观认知信息的情况下,借助客观信息(例如拓扑结构)来描述和分析.这个假设往往违背社会实践,例如一对异性关系中,可能一方认为双方是朋友,另一方却认为是恋人.社会关系的这种主观属性在分析人物主观行为(例如信息传播、好友选择等)时,较客观属性更加重要.

近年来,研究者们开始尝试在社会关系分析中引入互动信息.本文也将从互动语言内容信息入手,尝试对社会关系的主观性特征进行语义化的度量并进行可视化的分析.

### (I) 关系方向性研究

在方向性研究方面,利用3元闭包原理来导出未知关系或关系的未知方向是较为常见的方法,例如将3元闭包与其他3种社会学规则共同作为最优化方法的约束条件,来分析关系的方向<sup>[2]</sup>.除3元闭包外,更一般化的拓扑结构特征<sup>[3]</sup>、关系参与者的属性特征<sup>[4]</sup>以及内容特征<sup>[5]</sup>也被广泛用于关系方向的分析.

### (II) 关系强度研究

在强度研究方面,Granovetter<sup>[6]</sup>首先阐述了关系强度的概念,并指出了大量可用于分析强度的典型行为特征.近年来,基于互动行为信息的关系强度度量是较为主流的方法,例如将行为作为变量来度量关系强度<sup>[7]</sup>,或采用内容特征进行分析<sup>[8]</sup>等.此外,拓扑特征和3元闭包<sup>[9]</sup>以及内容信息也被用来分析关系强度.其中,部分工作<sup>[10]</sup>采用了互动中的标志性词汇作为特征之一,但未用到更丰富的语言元素.

### (III) 关系类型研究

有标记社会网络中的关系类型分析是近年来新的研究热点.目前,这方面工作主要关注关系的正负

极性分析,其中拓扑信息是最广泛的研究手段,3元闭包及平衡理论也得到应用.其中,Kunegis等<sup>[11]</sup>首先验证了正向边无法准确预测负向边,指出正负性研究的必要性.不同类型网络间的转移学习<sup>[12]</sup>和社会网络的谱分析方法<sup>[13]</sup>也是基于拓扑结构的关系类型研究的重要手段.除拓扑信息外,也有研究者将用户行为<sup>[14]</sup>和内容信息<sup>[15]</sup>作为辅助特征与平衡理论相结合进行关系类型分析.

社会语言学认为社会结构与语言结构互相影响.萨丕尔-沃尔夫假设<sup>[16]</sup>认为,语言结构决定着人们对世界的看法,即语言决定思维,进而思维决定行为与社会关系.

社会关系的内涵具有丰富的层次和内容,而传统的社会关系分析中所关注的方向、强度、极性和特定类型4种特征对关系内涵的描述能力是较为有限的.目前社会关系分析所广泛依赖的社会网络拓扑信息、行为信息和人物属性信息由于缺乏足够丰富的语义信息,难以刻画社会关系的丰富内涵.相比之下,互动语言内容所蕴含的丰富语义更有利于社会关系丰富内涵的描述,是社会关系画像研究的重要方向.

本文将尝试利用互动语言特征来语义化的度量在线社会关系的主观特征,并着重以可视化的方式展现并分析社会关系语义化度量的结果.

## 1 基于互动语言特征的社会关系语义化度量

### 1.1 社会语言学中的重要语言维度

语言的维度指的是社会语言学中用来分析语言的多个角度.社会语言的相关因素是多方面且复杂的,规定了有效的衡量维度之后,我们可以更准确更方便地通过对社会语言多个维度的度量来理解和分析社会中语言的使用.Holmes<sup>[17]</sup>对于研究语言在社会关系中的应用给出了4个重要维度:

(I) 亲疏社会的距离特征:关注在社会交往中的个体之间关系的远近.

(II) 社会地位的特征:关注在社会交往中个体之间的相对地位.

(III) 规范性的特征:关注在不同的关系、主题和场所下使用语言的规范性.

(IV)指示和情感的特征:关注在社会交往中语言的指示和情感功能.

其中,前两个维度关注的是主观和客观两方面的社会关系的特征.后两个维度关注的是互动语言的特性,并与社会关系高度相关.

### 1.2 互动语言特征的获取

本文可视化工作中的社会关系语义量化特征及计算沿用了前期工作中的方法<sup>[18]</sup>;我们选择了 4 种典型的互动语言特征来刻画个人的相互之间关系强度的主观观点,包括频率、长度、规范度和情感极性,分别表明了互动语言的数量、质量和情感特性.其中,频率和长度是语言交际中的两个基本特征,而规范度和情感极性相当于社会语言学中的规范性和情感特征.

本文以安然公司电子邮件数据集为例进行社会关系语义化度量方法的说明和展示.该数据集是 151 名安然雇员往来通信的电子邮件数据,作为政府调查安然公司财务所用数据的一部分.研究结果是基于 CMU 版(2015 年 5 月 7 日版)的集合,包含了约 0.5 M 的电子邮件数据.我们采用这个数据集是因为其中包含了时间戳、邮件内容、收发邮件的邮箱地址等信息,可以利用这些信息提取到这 4 个特征的值<sup>[17]</sup>.对于每一个定向的用户对  $\langle I_i, I_j \rangle$ ,我们通过对  $I_i$  发送给  $I_j$  的邮件内容进行抽取计算得到 4 个内容特征的值,并与  $I_j$  发送给  $I_i$  的邮件内容抽取到的 4 个特征值进行比较.

#### (I) 互动语言频率特征

在人际交往中,个体的语言互动频率包括了被动的回应他人和主动对与他人发起交流,可以部分地反映一个人在一段相互关系中的沟通意图.对于我们采用的邮件数据集来说,频率是指在一段特定时间内用户之间往来邮件的次数.

记  $I_i$  给  $I_j$  发送邮件的总数  $N$ ,从  $I_i$  到  $I_j$  发送的第一封邮件时间为  $t_1$ ,最后一封邮件之间的天数时间为  $t_2$ , $I_i$  对  $I_j$  “频率”特征值的计算公式为

$$\text{frequency\_score} = \frac{N}{t_2 - t_1} \quad (1)$$

#### (II) 互动语言长度特征

在人际交往中,互动语言的长度也是一个可以反映意图的指标,并且与个体对于双方之间的关系强度认知相关.本实验中,互动语言的长度指的是在用户间的往来邮件中的单词的平均数量.

记  $I_i$  给  $I_j$  发送的所有邮件的总词数  $W$ , $I_i$  给  $I_j$  发送的总邮件数为  $n$ , $I_i$  对  $I_j$  “长度”特征值的计算公式为

$$\text{length\_score} = \frac{W}{n} \quad (2)$$

#### (III) 互动语言规范度特征

在人际交往中,互动语言的规范度是一个可以反映说话者对这段相互关系重视程度的重要指标,这个重视程度是受个体对关系的主观认知而影响的.规范度可以反映交互语言中语言的形式和质量,与词语和语法的使用密切相关.

实验根据所有邮件中所有语句的混乱程度得分可以得到规范度,混乱程度越低则意味着语言的规范度越高,本文用 SRI 语言模型工具包(SRILM)来表达. $I_i$  对  $I_j$  的“规范度”特征值计算如下:

$$\text{perplexity\_score} = 10^{(-\log \text{prob}/(\text{words}-\text{oovs}+1))} \quad (3)$$

式中,prob 是句子出现的概率,words 和 oovs 分别为该句子中单词个数和 oovs 的个数.

#### (IV) 互动语言情感极性特征

情感极性是用来度量互动语言中情感倾向性的.我们利用一个情感词典来统计句子中情感词的个数,其中每个情感词权值为 1 或 -1,分别表示积极和消极的情感极性.然后根据用户在一段关系中发送邮件句子的平均情感极性,来计算该用户在对这段关系的情感极性.

记  $I_i$  给  $I_j$  发送的所有邮件中所有句子的情感词极性得分总和 Add, $I_i$  给  $I_j$  发送的所有邮件中的句子总数 total\_num,计算  $I_i$  对  $I_j$  “情感极性”特征值的公式为

$$\text{Sentiment\_score} = \frac{\text{Add}}{\text{total\_num}} \quad (4)$$

### 1.3 基于个人语言习惯的语言特征规范化

为了避免个人语言习惯的干扰,在本文的可视化方法中,我们采用了文献[19]中的规范化方法.对于个体  $I$ ,我们通过测量  $I$  的个人语言习惯的得分  $H_f(I)$ ,来规范化  $I$  的每一个语言特征  $f$ . $H_f(I)$  通过下式进行计算:

$$H_f(I) = \frac{1}{|C|} \sum_{I_i \in C} f(I, I_i) \quad (5)$$

式中, $f(I, I_i)$  是  $I$  对其他人  $I_i$  说的话的  $f$  值, $I_i$  属于  $C$ , $C$  是所有与  $I$  说过话的个体的集合.

$f(I, I_i)$  归一化后的值,通过下式进行计算:

$$f'(I, I_i) = \frac{f(I, I_i) - H_f(I)}{H_f(I)} \quad (6)$$

## 2 在线社会关系语义化度量的可视化与分析

基于上述研究,为了更加直观地展示和分析在

线社会关系特征的语义化度量结果,本文实现了一个面向在线社会关系语义化度量分析的可视化系统.该系统的主要功能如下:

(I) 社会网络全局分析:可视化展示出所有用户及用户间关系特征的语义化度量结果.

(II) 双用户局部网络对比分析:可视化表达任意用户对附近的局部社会关系结构及其语义化特征.其中社会网络以参比用户为中心.

(III) 3 用户结构平衡性分析:基于社会关系平衡理论,系统需要对于 3 用户构成的社会关系 3 元闭包的平衡进行局部和全局的可视化展示.

### 2.1 社会关系网络全局分析

算法 2.1 描述了社会关系网络全局分析的主要方法.该方法的主要思想是通过可视化元素来直观地展示整个网络的,尤其是关系边的量化语义特征.具体而言,首先以结点的直径来表达结点的度数(朋友数量);进而,对于任意一对相邻结点(朋友),用两条平行边来表达这一对结点间的双向语义特征值,值的大小则由边的宽度来实现可视化.

#### 算法 2.1 社会网络全局分析

输入:相互关系的集合  $E$ , 用户选择特征  $x$

输出:所有的用户和用户之间的所有相互关系,以及在特征  $x$  下的关系强度

初始化:  $V \leftarrow \emptyset, Ex \leftarrow \emptyset$

Foreach edge in  $E$  do

If  $e_{i,j} \notin E$  双边 then

$e.start \leftarrow edge.A;$

$e.end \leftarrow edge.B;$

$e.weight \leftarrow edge.e_{1x};$

$Ex \leftarrow e;$

$e.start \leftarrow edge.B;$

$e.end \leftarrow edge.A;$

$e.weight \leftarrow edge.e_{2x};$

$Ex \leftarrow e;$

Else

If  $edge.e_{1x} > edge.e_{2x}$  then

$e.start \leftarrow edge.A;$

$e.end \leftarrow edge.B;$

$e.weight \leftarrow edge.e_{1x} - edge.e_{2x};$

Else

$e.start \leftarrow edge.B;$

$e.end \leftarrow edge.A;$

$e.weight \leftarrow edge.e_{2x} - edge.e_{1x};$

If  $edge.A \notin V$  then

$V \leftarrow edge.A;$

If  $e_{i,j} \notin E$  单边 then

$V \leftarrow edge.B;$

$V \leftarrow edge.B;$

Return  $G \leftarrow (V, Ex)$

图 1~4 分别描述了采用 4 种不同语言特征进行社会关系语义化度量的情况下,社会网络全局的社会关系语义化度量分析.对于每一种特征,分别进行双向和无向两种度量.对于任一社会关系,双向度量指的是分别从关系参与双方的角度进行语义化度量,即分别利用关系参与双方各自的互动语言特征进行度量;而无向度量指的是将关系双方的双向度量合并为一个无向度量以展示该关系的整体度量结果.图 1~4 中,边的粗细程度表示相应社会关系语义度量值的大小,越粗表示值越大.而顶点的大小表示相应人物的社会关系数量,顶点越大表示社会关系越丰富.

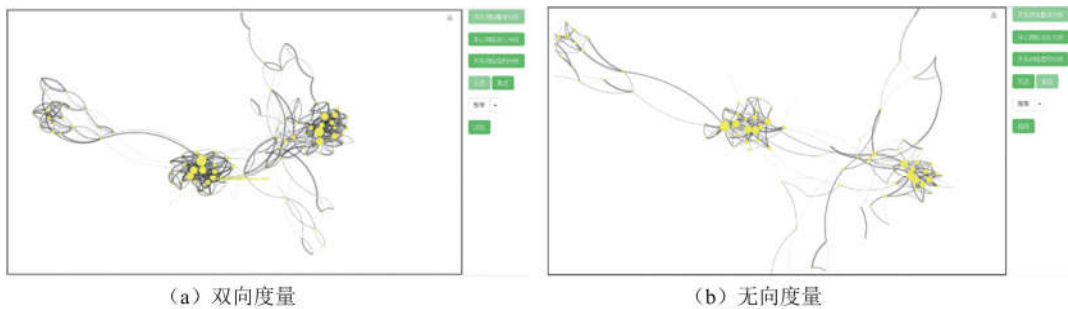


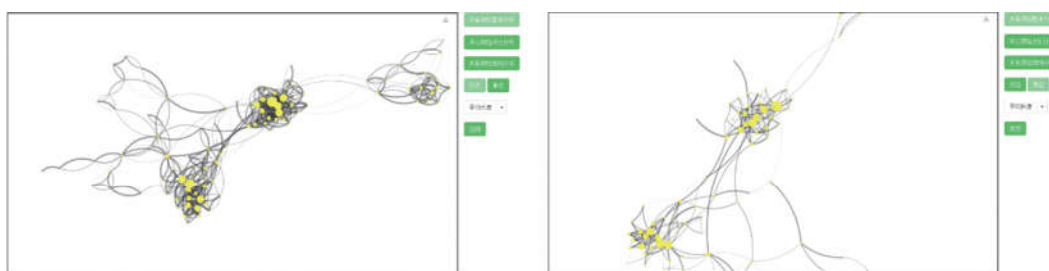
图 1 社会关系网络全局的语义化度量可视化(互动语言频率特征)

Fig.1 Visualization of global semantic measuring of social network with frequency feature

从图 1~4 可以看出,在无向度量中,不同社会关系的语义度量值有明显的区别,更重要的是,对于同一社会关系,关系双方的双向语义度量值也普遍存在明显差异.这说明社会关系参与双方对关系的主观认知常常存在显著差异,这种差异是传统的客

观性度量所难以刻画的.这种语义化度量的双向差异性在社会网络边缘较之网络中心更为显著.一种可能的解释是网络边缘人群可能是加入网络时间较短的个体(例如新进员工),因此在社交关系中会付出更多的单方面努力.同时,我们也发现,在一个社





(a) 双向度量

(b) 无向度量

图 2 社会关系网络全局的语义化度量可视化(互动语言长度特征)

Fig.2 Visualization of global semantic measuring of social network with length feature



(a) 双向度量

(b) 无向度量

图 3 社会关系网络全局的语义化度量可视化(互动语言规范度特征)

Fig.3 Visualization of global semantic measuring of social network with fluency feature



(a) 双向度量

(b) 无向度量

图 4 社会关系网络全局的语义化度量可视化(互动语言情感极性特征)

Fig.4 Visualization of global semantic measuring of social network with sentiment feature

会关系中,社会关系较少的一方总体而言比社会关系较多的一方更加重视双方之间的关系.

### 2.2 双用户局部网络对比分析

算法 2.2 描述了双用户中心网络对比分析的主要流程及方法.该方法的主要思想是在算法 2.1 全局网络可视化方法的基础上,允许用户选择两个不同的中心结点(人物)来进行比较,比较的目的是观察这两个结点周边的社会关系语义特征的异同.本方法进一步的用箭头来表示一对结点间的两条平行边所代表关系的方向.

#### 算法 2.2 双用户局部网络对比分析

输入:相互关系的集合 E,用户输入中心结点 C 和层数 l,以及特征 x

输出:以特定用户为中心结点指定关系层数的用户间相互关系及强度的网络图

```

初始化:  $V \leftarrow \emptyset, Ex \leftarrow \emptyset, T \leftarrow \emptyset, L \leftarrow \emptyset, EA \leftarrow \emptyset$ 
Foreach edge in E do
  If  $e_{i,j} \notin E$  双边 then
     $e.start \leftarrow edge.A;$ 
     $e.end \leftarrow edge.B;$ 
     $e.weight \leftarrow edge.e1x;$ 
   $Ex \leftarrow e;$ 
   $e.start \leftarrow edge.B;$ 
   $e.end \leftarrow edge.A;$ 
   $e.weight \leftarrow edge.e2x;$ 
   $Ex \leftarrow e;$ 
Else
  If  $edge.e1x > edge.e2x$  then
     $e.start \leftarrow edge.A;$ 
     $e.end \leftarrow edge.B;$ 
     $e.weight \leftarrow edge.e1x - edge.e2x;$ 

```

```

Else
e.start ← edge.B;
e.end ← edge.A;
e.weight ← edge.e2x - edge.e1x;
L[A] ← B;
L[B] ← A;
V ← C;
For i:0→1 do
Foreach v1 in V do
Foreach v2 in V do

```

```

If e[v1, v2] in Ex then
EA ← e[v1, v2];
V ← V ∪ L[v];
Return G ← (V, EA)

```

如图 5 所示,在双用户局部网络对比分析中,我们进一步发现,社会关系较少的节点较之社会关系较多的节点普遍具有更高的社会关系语义度量值(更粗的边),即更加重视其与他人的社会关系。

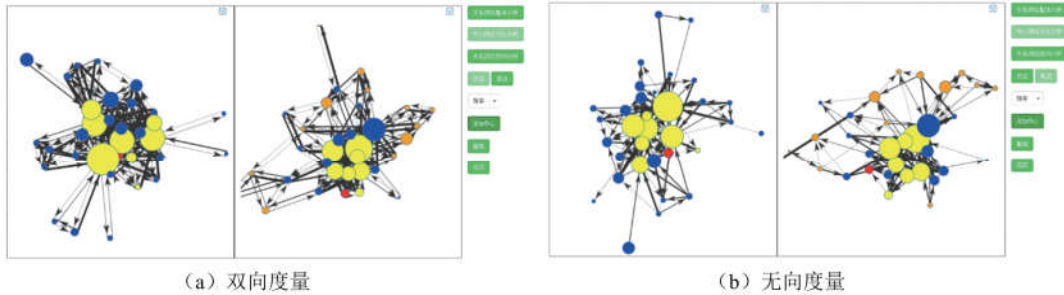


图 5 双用户附近局部网络的语义化度量可视化  
 Fig.5 Visualization of semantic measuring of user pairs' local social network

### 2.3 三用户结构平衡性分析

算法 2.3 描述了三用户结构平衡性分析的主要流程及方法.该方法的主要思想是在算法 2.1 的基础上,进一步强调网络中社会关系三元闭包结构的平衡性.方法利用色彩突出标识网络中的平衡三角形,从而可以直观地观察整个网络的平衡情况,从而估计一个社会网络的稳定程度。

#### 算法 2.3 三用户结构平衡性分析

输入:相互关系中所有可以构成三角形的关系的集合 E,用户选择特征 x

输出:所有的用户和用户之间的所有相互关系及强度,并标注符合平衡理论的三角形

初始化:  $V \leftarrow \emptyset, Tx1 \leftarrow \emptyset, Tx2 \leftarrow \emptyset, Tx3 \leftarrow \emptyset, dif1 = 0, dif2 = 0, dif3 = 0$

```

Foreach tri in E do
dif1 = dif1 + tri.e1x + tri.e2x;
dif2 = dif2 + |tri.e1x - tri.e2x|;
dif3 = dif3 + |tri.e3x - tri.e4x|;
If tri.A ∉ V then
V ← tri.A;
If tri.B ∉ V then
V ← tri.B;
If tri.C ∉ V then
V ← tri.C;
dif1 = dif1/E.size;
dif2 = dif2/E.size;
dif3 = dif3/E.size;
Foreach tri in E do

```

```

tri.color = black;
If tri.e1x + tri.e2x > dif1 then
tri.flag = tri.flag + 1
If |tri.e3x - tri.e4x| < dif3 then
tri.color = red;
Tx3 ← tri;
Else
If |tri.e3x - tri.e4x| > dif3 then
tri.color = red;
Tx3 ← tri;
If |tri.e1x - tri.e2x| > dif2 then
If |tri.e3x - tri.e4x| > dif3 then
tri.color = red;
Tx2 ← tri;
Else
If |tri.e3x - tri.e4x| < dif3 then
tri.color = red;
Tx2 ← tri;
Foreach tri in E do
If tri.flag == 1 or tri.flag == 3 then
tri.color = red;
Tx1 ← tri;
If ei,j ∉ E 传统平衡理论 then
Return G ← (V, Tx1)
Elif ei,j ∉ E 基于不对称度量的平衡理论
差值变体 then
Return G ← (V, Tx2)
Else 基于不对称度量的平衡理论加和

```

变体 then

Return  $G \leftarrow (V, T \times 3)$

图 6 和 7 展示了三用户结构平衡性分析的可视化结果.其中边的粗细仍然表示语义化度量值的大小.深灰色三角形表示平衡的社会关系 3 元闭包,而

黑色三角形则表示不平衡的社会关系 3 元闭包.从图 6 和 7 可以看到,无论对于经典还是扩展的 3 元闭包理论,语义化度量结果都符合理论预测,即社会网络中多数社会关系 3 元闭包倾向于平衡.



图 6 基于经典三元闭包平衡理论的三用户社会关系语义化度量的平衡性分析可视化

Fig.6 Visualization of the balance of triple users' semantic relationship measuring with traditional balance theory



图 7 基于扩展的三元闭包平衡理论的三用户社会关系语义化度量的平衡性分析可视化

Fig.7 Visualization of the balance of triple users' semantic relationship measuring with extended balance theory

## 2.4 讨论与分析

从上述结果中可以看到,基于可视化方法,我们可以更加直观地观察到社会关系语义化度量方法所展示的社会关系网络中个体、局部乃至全局的社会关系语义分布化分布特征.

通过上述结果,我们还可以看到,社会关系语义化度量及其可视化分析方法与现有社会关系分析方法相比较,主要的新意在于:

(I) 主观性假设: 现有研究多将社会关系的分析视为客观任务,而在许多场景中,关系双方对关系的主观认知更加重要.本文将人际互动语言内容作为主观性信息的来源,并以此来刻画关系参与者对关系内涵的主观认知.

(II) 语义化表达: 基于社会直接采用互动语言内容的多种语言特征来多维度地描述社会关系的语义化特征.互动语言与目前社会关系分析常用的人物自然社会属性、拓扑特征乃至个人言论特征相比,都具有更加丰富的语义信息,并且与人际互动关系的语义内涵具有更加直接的关系.

(III) 双向主观性度量: 将社会关系分析的双向

性建立在社会关系的主观性上.换言之,正是因为社会关系双方对于社会关系的主观看认知是不对称的,所以社会关系度量的双向性便成为自然而必要的.

(IV) 基于可视化的社会意义观察: 社会关系语义化度量的可视化是本研究的核心内容.与具有客观评测数据的传统社会关系分析方法相比,在暂时缺乏可用的客观评测数据,难以进行定量分析的情况下,通过可视化,我们可以直观地观察到社会关系语义化度量在个体、局部和全局的分布特征,从而获得直观的启发式结果,并将之与经验相比较,来初步分析社会关系语义化度量的合理性.

## 3 结论

本文通过设计并实现了一套可视化方法来直观地展示和分析社会关系的语义化主观特征,实现了社会关系语义化度量的可视化分析.具体包括了社会网络全局分析、双用户局部网络对比分析和 3 用户结构平衡性分析.在可视化分析中,以电子邮件社会网络为例,我们实现了双向和无向语义化度量值的可视

化,并从全局、局部和三元闭包三个主要角度观察了社会网络中社会关系的语义化特征所具有的若干特点,其中包括社会网络中社会关系双向语义度量值的差异、用户社会关系系数与双向语义度量值差异之间的关系以及社会网络中三元闭包的总体平衡性等。

#### 参考文献(References)

- [1] 王飞跃,李晓晨,毛吉文,等. 社会计算的基本方法与应用[M]. 2版,杭州:浙江大学出版社,2013.
- [2] ZHANG J, WANG C K, WANG J M. Who proposed the relationship? Recovering the hidden directions of undirected social networks[C]// Proceedings of the 23rd International Conference on World Wide Web. Seoul; ACM Press, 2014: 807-817.
- [3] CHIANG K Y, NATARAJAN N, TEWARI A. Exploiting longer cycles for link prediction in signed networks [C]// Proceedings of the 20th ACM International Conference on Information and knowledge management. Glasgow, UK; ACM Press, 2011: 1157-1162.
- [4] LESKOVEC J. How users evaluate each other in social media[C]// Proceedings of the Web Search Data Mining. Seattle; ACM Press, 2012.
- [5] YUAN N J, ZHONG Y, ZHANG F Z, et al. Who will reply to/retweet this tweet? The dynamics of friendships and online social interactions [C]// Proceedings of the Web Search Data Mining. San Francisco; ACM Press, 2016: 3-12.
- [6] GRANOVETTER M S. The strength of weak ties[J]. The American Journal of Sociology, 1973, 78(6): 1360-1380.
- [7] JIONGQIAN LIANG, DEEPAK AJWANI, PATRICK NICHOLSON, et al. What links Alice and Bob? Matching and ranking semantic patterns in heterogeneous networks [C]// Proceedings of the 25th International Conference on World Wide Web. Québec, Canada; IEEE Press, 2016: 879-889.
- [8] ZHUANG J F, MEI T, HOI S C H, et al. Modeling social strength in social media community via kernel-based learning [C]// Proceedings of the 19th ACM international conference on Multimedia. Scottsdale; ACM Press, 2011: 113-122.
- [9] SINTOS S, TSAPARAS P. Using strong triadic closure to characterize ties in social networks [C]// Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining. New York; ACM Press, 2014: 1466-1475.
- [10] ADALI S, SISENDA F, MAGDON-ISMAIL M. Actions speak as loud as words; Predicting relationships from social behavior data [C]// Proceedings of the 21rd International Conference on World Wide Web. Lyon, France; ACM Press, 2012: 689-698.
- [11] KUNEGIS J, PREUSSE J, SCHWAGEREIT F. What is the added value of negative links in online social networks? [C]// Proceedings of the 22rd International Conference on World Wide Web. Rio de Janeiro, Brazil; ACM Press, 2013: 727-736.
- [12] YE J H, CHENG H, ZHU Z, et al. Predicting positive and negative links in signed social networks by transfer learning [C]// Proceedings of the 22rd International Conference on World Wide Web. Rio de Janeiro, Brazil; ACM Press, 2013: 1477-1488.
- [13] KUNEGIS J, SCHMIDT S, LOMMATZSCH A. Spectral analysis of signed graphs for clustering, prediction and visualization [C]// Proceedings of the SIAM International Conference on Data Mining. Columbus, USA; ACM Press, 2010: 559-559.
- [14] TANG J L, CHANG S Y, AGGARWAL C, et al. Negative link prediction in social media [C]// Proceedings of the 8th ACM International Conference on Web Search and Data Mining. Shanghai, China; ACM Press, 2015: 87-96.
- [15] WESTR, PASKOV H S, LESKOVEC J, et al. Exploiting social network structure for person-to-person sentiment analysis[J]. IEEE Transactions of the Association for Computational Linguistics, 2014; arXiv:1409.2450.
- [16] SAPIR E. The status of linguistics as a science[J]. Linguistic Society of America, 1929, 5(4): 207-214.
- [17] HOLMES J. An Introduction to Sociolinguistics[M]. 4ed, Rutledge, 2013.
- [18] WANG B, YU Y S, ZHANG P. Investigation of the subjective asymmetry of social interrelationship with interactive language [C]// Proceedings of the 25rd International Conference on World Wide Web. Geneva, Switzerland; ACM Press, 2016: 121-122.
- [19] WANG B, SUN Y J, HAN B, et al. Extending the balance theory by measuring bidirectional opinions with interactive language [C]// Proceedings of the 26rd International Conference on World Wide Web. Perth, Australia; ACM Press, 2017: 849-850.