RoleNet: Movie Analysis from the Perspective of Social Networks

Chung-Yi Weng, Wei-Ta Chu, Member, IEEE,
and Ja-Ling Wu, Fellow, IEEE

IEEE Transaction on Multimedia Vol. 11 No.2
Feb,2009

Speaker: Yi-Lin , Hsu
Advisor: Dr. Koh, Jia-ling   Date: 03/08/2010

Introduction

- Over the past decade, researches on movie analysis attempt to solve the most notorious problem—the semantic gap. However, it seems that approaches based on audiovisual features face an unbreakable impediment.
- These studies come from "frame-level" analysis, which is based on shot change detection and keyframe selection to "event-level" analysis, which further considers the temporal context or objects in the scenes and achieves the detection of some important events such as dialog and gunplay .

Introduction

- In this work, we propose a story-level analysis system based on the social relationships between characters.
- We proposed an SNA-based approach to analyze movies in . Leading roles and corresponding communities can be automatically identified by checking the social relationships between characters.

Definition of RoleNet

- A model that is suitable to describe roles' relationship should possess the following characteristics.
  - Representing relationships effectively
  - Facilitating systematic analysis:
- *Definition: A RoleNet is a weighted graph expressed by*

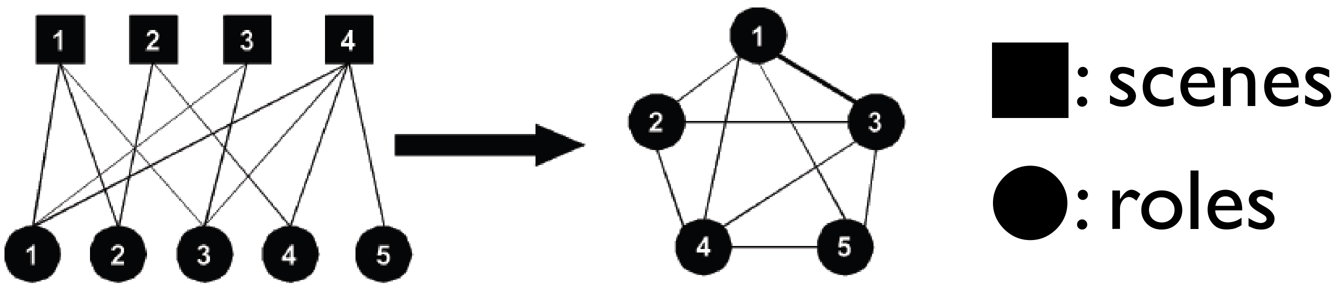
$$G = \langle V, E, W \rangle$$

- *Where* $V = \{v_1, v_2, \ldots, v_n\}$

$$E = \{e_{ij} | \text{if } v_i \text{ and } v_j \text{ have relationship}\}$$

W : w_{ij} in W represents the strenth of the relationship between v_i and v_j

Construction of RoleNet



$$a_i^T a_j = w_{ij} \text{ for } i \neq j$$

$$w_{ii} = 0$$

$$A_{m\times n} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 \end{bmatrix} \longrightarrow W_{n\times n} = \begin{bmatrix} 0 & 1 & 3 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 3 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \end{bmatrix}$$

$$A^T A = W$$

Community Analysis

- RoleNet Construction
- Leading roles determination
- Community identification

Community Analysis –Bilateral movie analysis

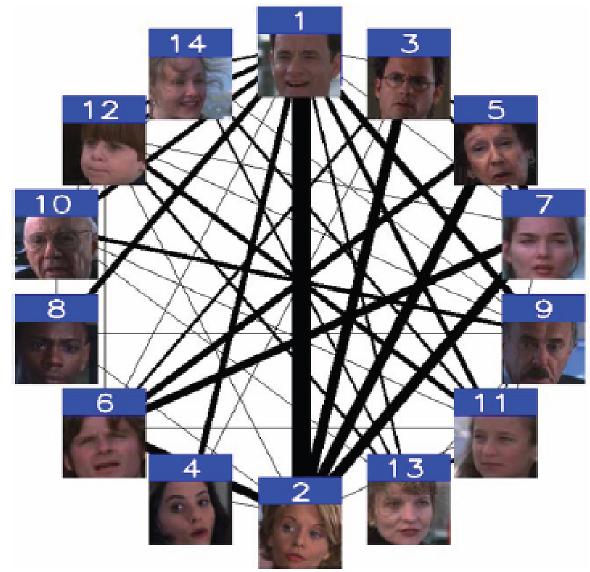


TABLE I
ROLES IN THE MOVIE "YOU'VE GOT MAIL"

Node (Role)	Meaning of roles
1	The hero (Tom Hanks)
2	The heroine (Meg Ryan)
3, 5, 6, 7	The heroine's friends and colleagues
4, 8, 9, 10, 11, 12, 13, 14	The hero's friends, relatives, and colleagues.

Leading roles determination

- In SNA, evaluating the impact of each individual is one of the earliest issues. It is known as the *centrality problem*.
- Based on RoleNet, we evaluate the centrality of the node (role) as

$$c_i = \sum_{j \neq 1} w_{ij}$$

Community identification

Given a RoleNet, find a labeling solution Δ^*:

$$\Delta^* = \arg\min_{\Delta} C(\Delta) \text{ subject to } \delta_p = 0 \text{ and } \delta_q = 1 \quad (5)$$

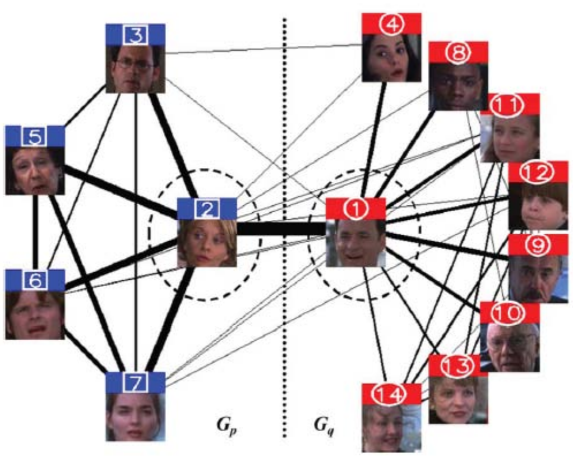
$$C(\Delta) = \sum_{i,j} |\delta_i - \delta_j| w_{ij} \quad (6)$$

$$\Delta = \{\delta_i, i = 1, \ldots, n\} \quad (7)$$

- Vp: the first leading role , Vq: the second leading role
- △is a set of binary labels :
  δi= 0 , if vi is assigned to the community led by vp
  δi= 1 , if vi is assigned to the community led by vq
- C(△) Is the closeness between two communities.

Community Analysis –Generalization

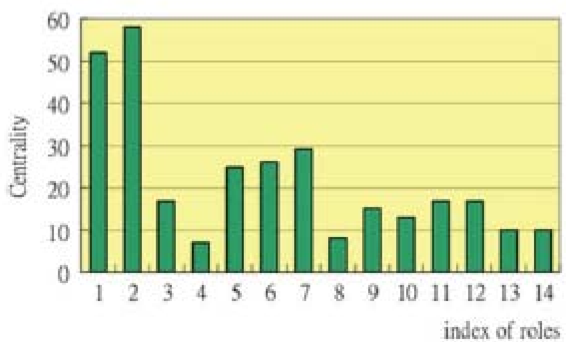
- Automatically determining the number of leading roles
- Analyzing finer communities
  - Micro (i.e. 4 and 8)

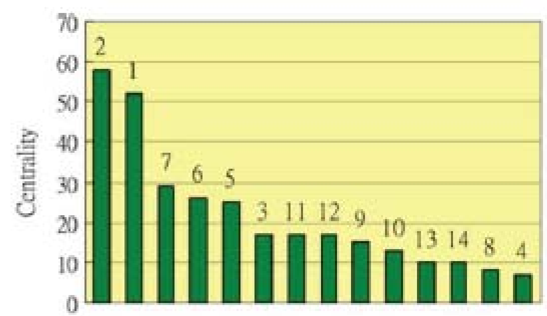


Macro	Micro	Meaning of roles
1		The hero (Tom Hanks)
2		The heroine (Meg Ryan)
3, 5, 6, 7	3	The heroine's boy friend
	5, 6, 7	The heroine's colleagues
4, 8, 9, 10, 11, 12, 13, 14	4	The hero's girl friend
	8	The hero's assistant
	9, 10	The hero's father and grandfather. The hero and they are co-founders of a company.
	11, 12	The hero's niece and nephew. They just visit the hero at holiday.
	13, 14	The hero's stepmother and her servant

Leading roles determination

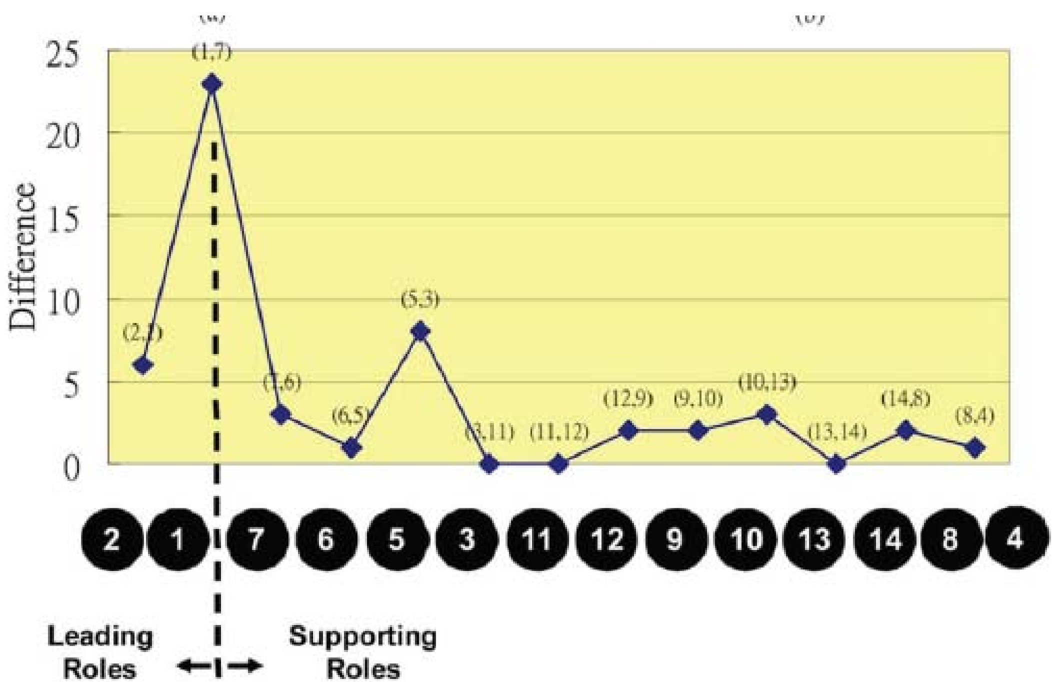
- Calculate the centrality value of each role
- Sort the centrality values in descending order





Leading roles determination

- Calculate the centrality difference between two adjacent roles.
- Find the maximum point in the difference distribution.



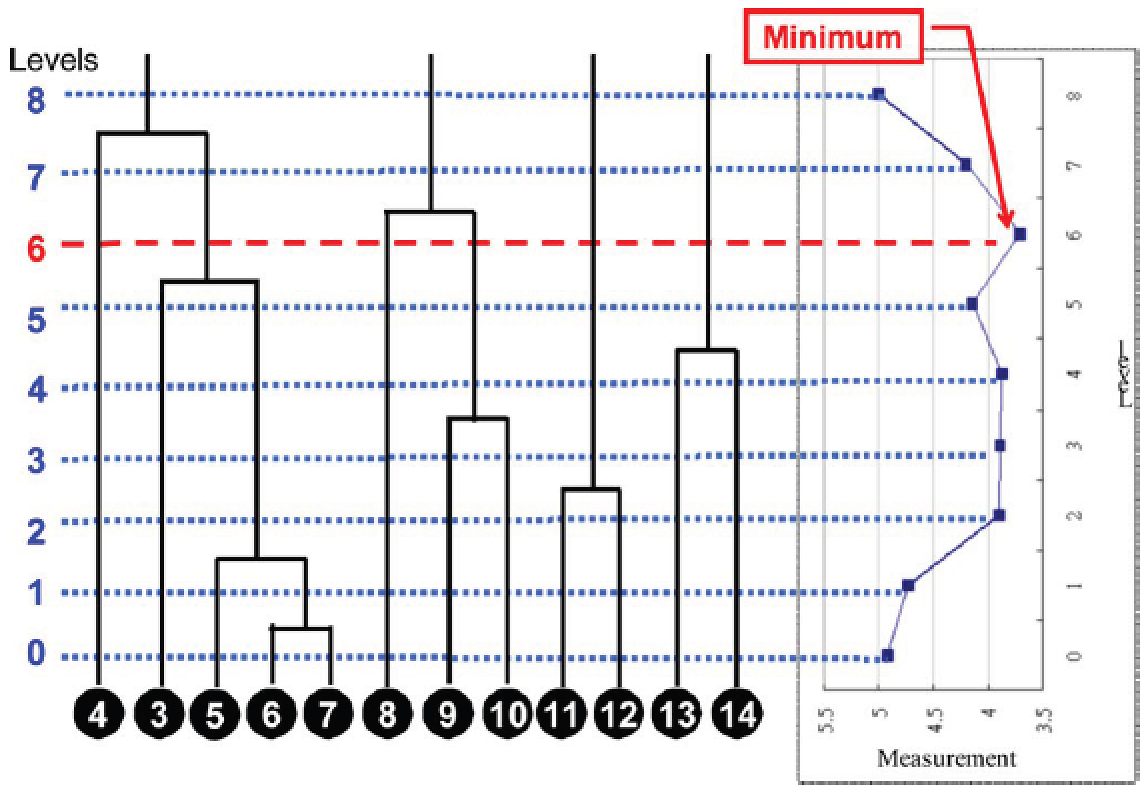
Micro-Community Identification

- Remove the leading roles and the edges linked to them from the RoleNet.

Algorithm 2: Micro-Community Identification

1. Initialize every individual node as a micro-community. The set of micro-community is denoted as $\Pi_t = \{T_1^t, T_2^t, \ldots, T_n^t\}, t = 0$, if there are initially n individual nodes. The size of the pth community in Π_t is denotes as $|T_p^t|$, which is the number of nodes included in this community.
2. From the modified RoleNet, find the edge that has the largest weight, say the edge e_{ij} between the node v_i and the node v_j, $v_i \in T_p^t$ and $v_j \in T_q^t$, then
   1) If $|T_p^t| \geq 1$ and $|T_q^t| = 1$, then $T_p^{t+1} = T_p^t \cup T_q^t, \Pi_{t+1} = \Pi_t - \{T_q^t\}$, and $t = t + 1$.
   2) If $|T_p^t| > 1$ and $|T_q^t| > 1$, then keep current community situation.
3. Remove the edge e_{ij} from the modified RoleNet and go to *Step* 2 until all edges have been removed.

Micro-Community Identification



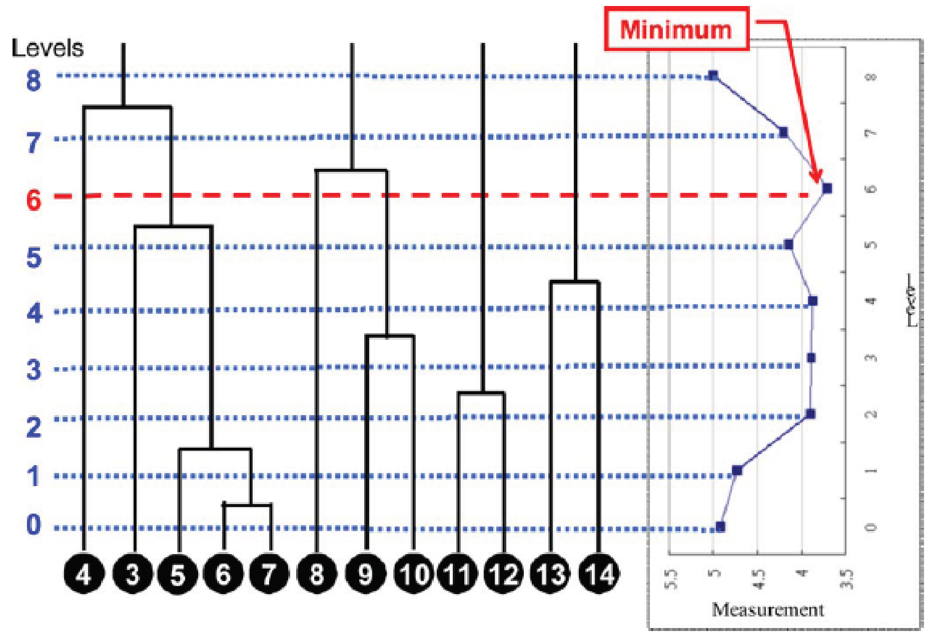
Micro-Community Identification

- We design a measurement to evaluate the community case at different levels. For the level t , the measurement is defined as

$$AvgW_t = \frac{\sum w_{ij}}{||\Pi_t||}, \forall v_i \in T_p^t, v_j \in T_q^t, p \neq q$$

Macro-Community Identification

- {4} ,
  {3,5,6,7},
  {8},
  {9,10},
  {11,12},
  {13,14}

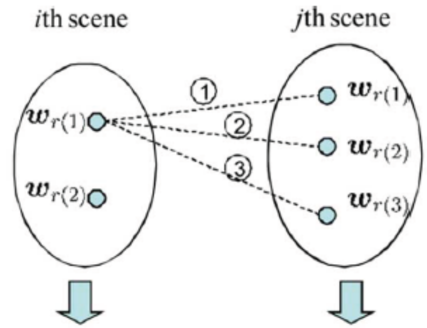


$$v^* = \arg\max_{v_i \in L}(\max_{v_j \in T_p} w_{ij})$$

Story Segmentation

- Scene Representation
  - We describe scenes by "the context of roles" rather than audiovisual features. Story segmentation is achieved by comparing the role's context in successive scenes.
- Story Segmentation

Scene Representation



$$CM_i = \begin{bmatrix} \boldsymbol{w}_{r(1)} & \boldsymbol{w}_{r(2)} \end{bmatrix} \quad CM_j = \begin{bmatrix} \boldsymbol{w}_{r(1)} & \boldsymbol{w}_{r(2)} & \boldsymbol{w}_{r(3)} \end{bmatrix}$$

Context-based similarity:

$$CM_i^T CM_j = \begin{array}{c} \begin{array}{ccc} ① & ② & ③ \end{array} \\ \begin{bmatrix} \boldsymbol{w}_{r(1)}^T \boldsymbol{w}_{r(1)} & \boldsymbol{w}_{r(1)}^T \boldsymbol{w}_{r(2)} & \boldsymbol{w}_{r(1)}^T \boldsymbol{w}_{r(3)} \\ \boldsymbol{w}_{r(2)}^T \boldsymbol{w}_{r(1)} & \boldsymbol{w}_{r(2)}^T \boldsymbol{w}_{r(2)} & \boldsymbol{w}_{r(2)}^T \boldsymbol{w}_{r(3)} \end{bmatrix} \end{array}$$

Fig. 9. Example of calculating the context-based similarity between two scenes.

$$\boldsymbol{w}_{r(k)} = (w_{1r(k)}, w_{2r(k)}, \ldots, w_{nr(k)}) \quad \text{(Normalized)}$$

$$d_{ij} = 1 - \frac{1}{pq} \sum_{s=1}^{p} \sum_{t=1}^{q} CM_{ij}(s,t)$$

Story Segmentation

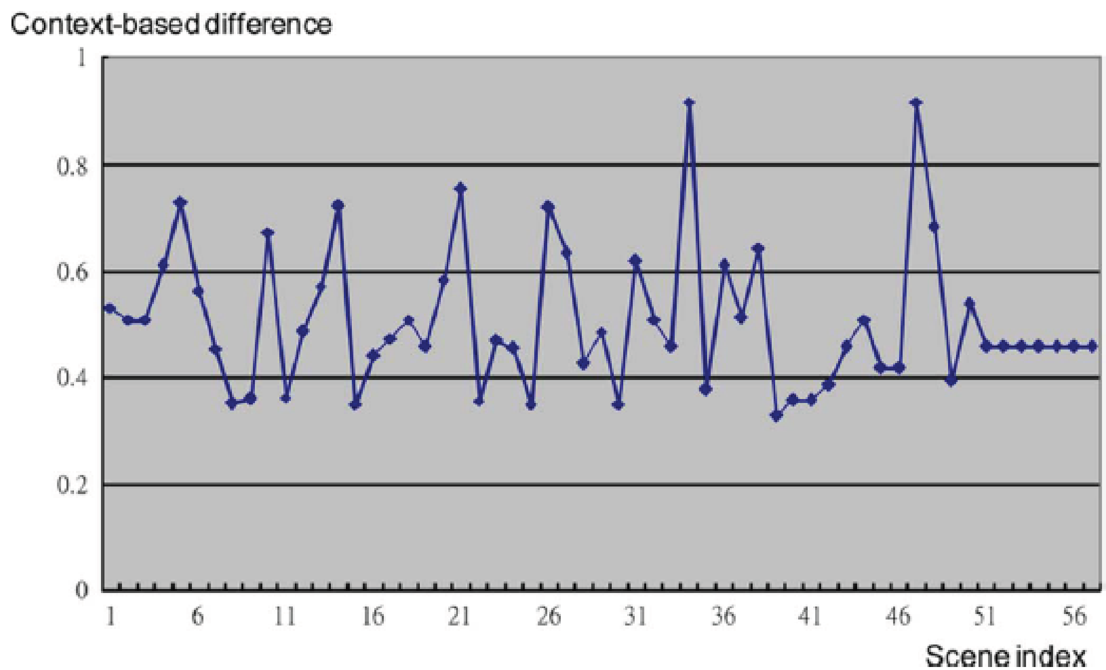


Fig. 10. Context-based difference curve for the movie "You've Got Mail".

- 2<= i <=N-2

$$\begin{cases} b_i \in Y, & \text{if } d_i < d_{i-\alpha 1} \text{ and } d_i < d_{i+\alpha 2} \\ b_i \in P, & \text{if } d_i > d_{i-\alpha 1} \text{ and } d_i > d_{i+\alpha 2} \\ b_i \in OT, & \text{otherwise} \end{cases}$$

$$\alpha_1 = \min\{j | j \in A_1\}$$

$$\alpha_2 = \min\{j | j \in A_2\}$$

$$A_1 = \{k | (d_i - d_{i-k}) \neq 0, 1 \leq k \leq i-1\}$$

$$A_2 = \{k | (d_i - d_{i+k}) \neq 0, 1 \leq k \leq (N-1-i)\}$$

Story Segmentation

Algorithm 3: The Storyshed Algorithm

Input: The set of scene boundaries $B = \{b_1, b_2, \ldots, b_{N-1}\}$ and the corresponding context-based difference values $D = \{d_1, d_2, \ldots, d_{N-1}\}$.
Output: The set of story boundaries.

1. According to D, find the boundaries in the valley set Y and the peak set P.
2. For each valley in Y, find the two nearest peaks from P that are respectively at the left and the right of it.
3. For each valley y_i in Y, fill water into each valley until the height of the water horizontal just floods one of the corresponding peaks.
4. Pick the scene boundaries b_k which have context-based difference values no less than the water horizontal and are located between y_i's two peaks. The set of story boundaries $SB = SB \cup \{b_k\}$.

Story Segmentation

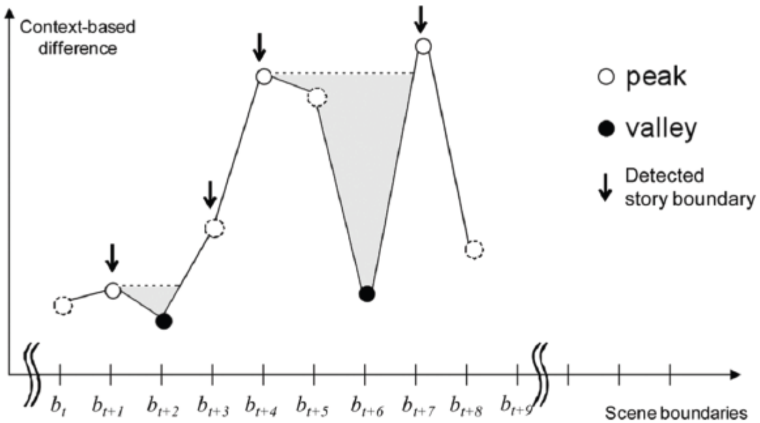


Fig. 11. Example of the storyshed segmentation method without global threshold.

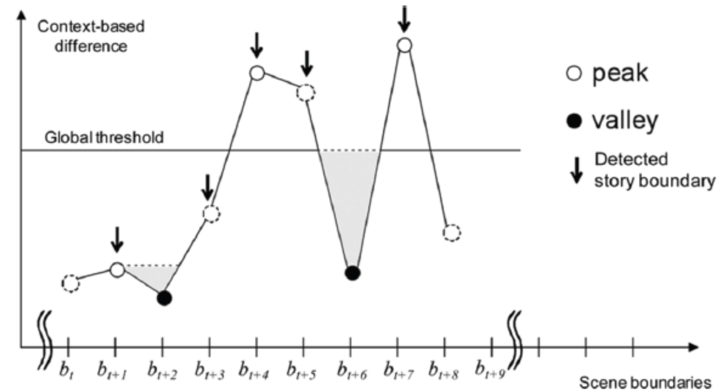


Fig. 12. Example of the storyshed segmentation method with a global threshold.

Evaluation

- DataSet
  - 10 Hollywood movies and 3 TV shows to evaluate the methods.
    - These movies have different numbers of leading roles.
  - The total length of the evaluation is over 21 hr and 428 story segments are included.
- Over 97% of scenes actually contain face information, which provides a solid foundation for us to reveal role's social relationship.
- The experimental results are based on the manually-labeled data.
- Automate the whole process and demonstrate the corresponding results.

Community Analysis

TABLE IV
DIFFERENT SITUATIONS IN DEFINING THE COMMUNITY GROUND TRUTH

Subject A's opinion	Subject B's opinion	Subject C's opinion	Final ground truth
{1,2,3} {4,5,6}	{1,2,3} {4,5,6}	{1,2,3} {4,5,6}	{1,2,3} {4,5,6}
{1,2,3} {4,5,6}	{1,2,3} {4,5,6}	{1,2,3,4} {5,6}	{1,2,3} {4,5,6}
{1,2,3} {4,5,6}	{1,2} {3,4} {5,6}	{1,2,3,4} {5,6}	Anyone
{1,2,3} {4,5,6}	{1} {2,3} {4,5,6}	{1,2,3,4} {5,6}	Anyone

TABLE V
PERFORMANCE OF LEADING ROLE DETERMINATION AND MACRO-COMMUNITY IDENTIFICATION

Movie ID	Ground truth	Determined leading roles	# of roles categorized correctly / # of roles
M1	1	1	12 / 12
M2	1, 2	1, 2	14 / 14
M3	1, 2, 6	1, 2, 6	20 / 20
M4	1, 2	1, 2	15 / 15
M5	1, 2	1, 2	8 / 9
M6	1	1	9 / 9
M7	1	1	15 / 15
M8	1	1	15 / 15
M9	1	1	15 / 15
M10	1	1	10 / 10
S1	1	1	14 / 14
S2	1, 2, 4, 5, 6, 7	1, 2, 3, 4, 5, 6, 7, 9	10 / 12
S3	1, 2, 3, 4	1, 2, 3	7 / 12

Performance of Leading Roles Determination

- The promising performance comes from two reasons.
  - 1) Leading roles pass through most scenes in a movie and have close relationship with others.
  - 2) Based on the representation of RoleNet, leading roles can be clearly identified by measuring the impact of different roles.
- The performance in TV shows is worse than that in movies.
  - TV shows often last for less than 30 minutes
  - TV shows have fewer than 30 scenes.
  - The pace of shows is fast, because directors have to use short and fewer scenes to present stories.
  - The selected TV shows include many leading roles.
- People can infer what happen and understand the subtle relationships between roles quickly, but the proposed method still appeals to the well-constructed relationships based on the frequent co-occurrence of roles.

Performance of Community Identification

- The performance of the proposed community process is very promising for movies.
- The trend of mutual relationship is apparent, and the proposed method catches this characteristic.
- The identification performance of TV shows is worse because we face the same situation as described in the previous section.
- To verify the length issue, we especially concatenate two episodes of "Sex and The City" (Season 2, Episodes 11 and 12) into a one-hour video and perform the same processes for leading role determination and community analysis.
- All four leading roles are correctly determined, and the results of community analysis are much better than that of analyzing one episode only.
- This result verifies the length issue and reveals the limitation of the proposed methods as well.

TABLE VI

DETAILED PERFORMANCE OF MICRO-COMMUNITY IDENTIFICATION

Movie ID	Ground truth	Results of micro-community identification
M1	Leading roles:{1} Other Roles: {2,3,4,9,11},{5,6,7}, {8},{10},{12}	{2,3,4,9},{5,6,7},{8}, {10},{11},{12}
M2	Leading roles:{1,2} Other Roles: {3},{4},{5,6,7},{8}, {9,10},{11,12},{13,14}	{3,5,6,7},{4},{8}, {9,10},{11,12},{13,14}
M3	Leading roles:{1,2,6} Other Roles: {3,4,5},{7},{10,11,12,13,16},{8, 9,18}, {14,17}, {15},{19},{20}	{3,4,5},{7,10,11,12,13,16} , {8,9,18},{14,17},{15},{19 },{20}

$$R = \frac{\sum_{i=1}^{k} \sum_{j \neq i} \delta_{ij}}{2 \times \binom{k}{2}}$$

$$\begin{cases} \delta_{ij} = 1, & \text{if}, \zeta_{ij}^{g} = 1 \text{ and } \zeta_{ij}^{v} = 1 \\ \delta_{ij} = 1, & \text{if}, \zeta_{ij}^{g} = 0 \text{ and } \zeta_{ij}^{v} = 0 \\ \delta_{ij} = 0, & \text{otherwise} \end{cases}$$

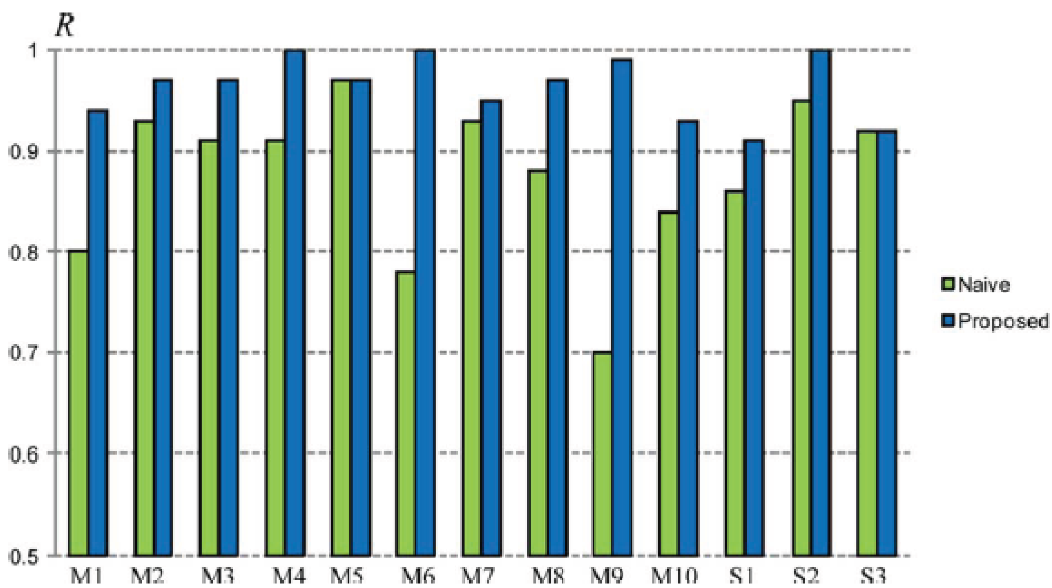
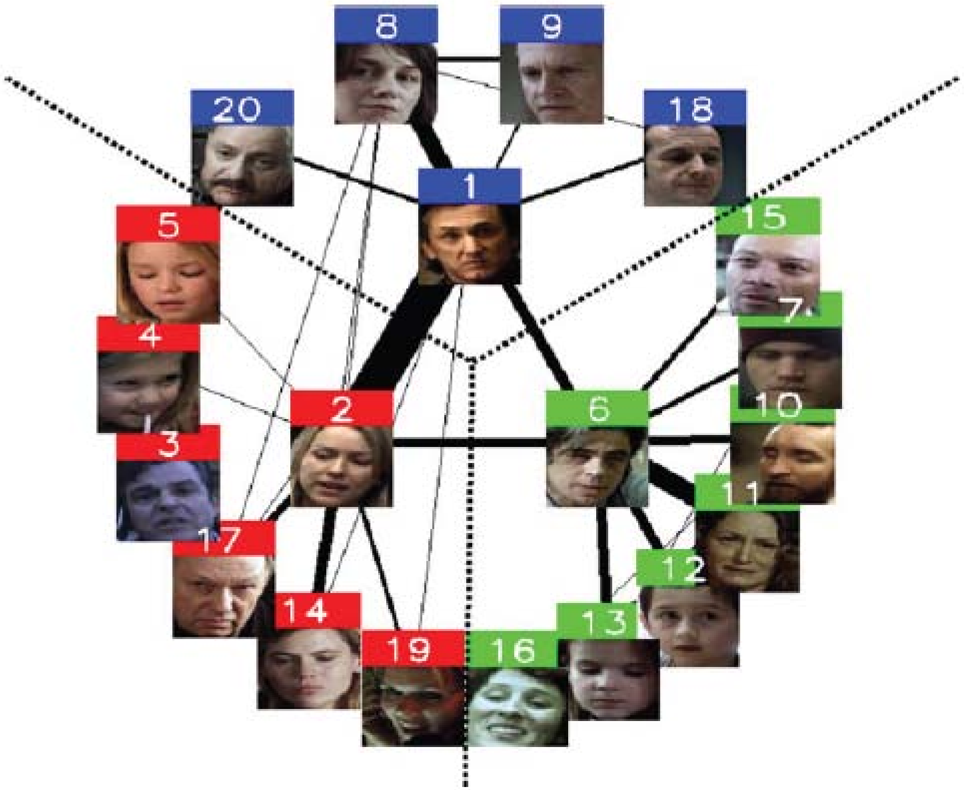


Fig. 13. Performance comparison of micro-community identification based on the proposed quantification method.


8
9
20
18
1
5
15
7
4
2
6
10
3
11
17
12
14
13
19
16

Purity

Ground Truth

Experiment

$$\rho = \left(\sum_{i=1}^{N_g} \frac{\tau(s_i)}{T} \sum_{j=1}^{N_v} \frac{\tau^2(s_i, s_j^*)}{\tau^2(s_i)} \right) \cdot \left(\sum_{j=1}^{N_v} \frac{\tau(s_j^*)}{T} \sum_{i=1}^{N_g} \frac{\tau^2(s_i, s_j^*)}{\tau^2(s_j)} \right)$$

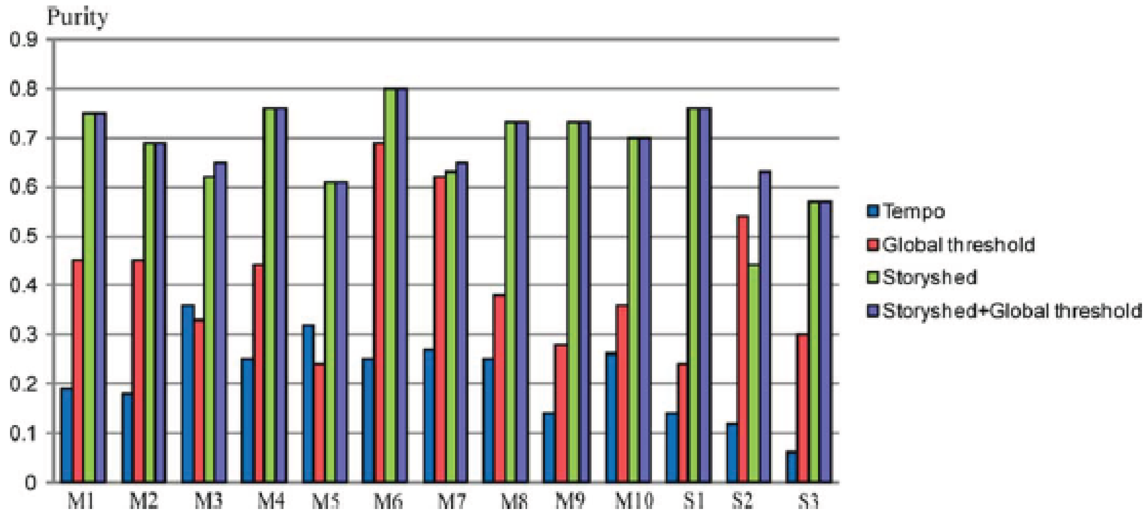


TABLE VII
OVERALL PURITY OF STORY SEGMENTATION IN DIFFERENT APPROACHES

	Tempo	Global threshold	Storyshed	Storyshed + Global threshold
Overall	0.21	0.41	0.68	0.69

- H.-W. Chen, J.-H. Kuo, W.-T. Chu, and J.-L.Wu, “Action movies segmentation and summarization based on tempo analysis,” in *Proc. ACM SIGMM Int.Workshop on Multimedia Information Retrieval, 2004, pp.* 251–258.

Conclusion

- The idea of SNA to do movie analysis
- An approach to model roles' interrelationship as a network
- Novel algorithms to analyze social relationships in movies
- A social-relation-based story segmentation method