# Michigan at DUC04 – Part I
# Using sentence prestige for document summarization

Güneş Erkan[1] and Dragomir Radev[1,2]

[1]Department of EECS
[2]School of Information

# Introduction

- Tasks 1-5
- Tasks 3-4 in this talk
- Task 5 (tomorrow, Güneş Erkan)

# Social networks

- Induced by a relation
- Allison and Bill are friends
- Prestige (centrality) in social networks:
  - Degree centrality: number of friends
  - Geodesic centrality: bridge quality
  - Eigenvector centrality: who your friends are
- Recommendation systems

# Eigenvectors of stochastic graphs

- Square connectivity matrix
- Directed vs. undirected
- An eigenvalue for a square matrix $A$ is a scalar $\lambda$ such that there exists a vector $x \neq 0$ such that $Ax = \lambda x$
- The normalized eigenvector associated with the largest $\lambda$ is called the principal eigenvector of $A$
- A matrix is called a stochastic matrix when the sum of entries in each row sum to 1 and none is negative. All stochastic matrices have a principal eigenvector
- The connectivity matrix used in PageRank [Page & al. 1998] is irreducible [Langville & Meyer 2003]
- An iterative method (power method) can be used to compute the principal eigenvector
- That eigenvector corresponds to the stationary value of the Markov stochastic process described by the connectivity matrix
- This is also equivalent to performing a random walk on the matrix

# Eigenvectors of stochastic graphs

- The stationary value of the Markov stochastic matrix can be computed using an iterative power method:

$$p = E^T p$$

$$(I - E^T) p = 0$$

- PageRank adds an extra twist to deal with dead-end pages. With a probability 1-$\varepsilon$, a random starting point is chosen. This has a natural interpretation in the case of Web page ranking

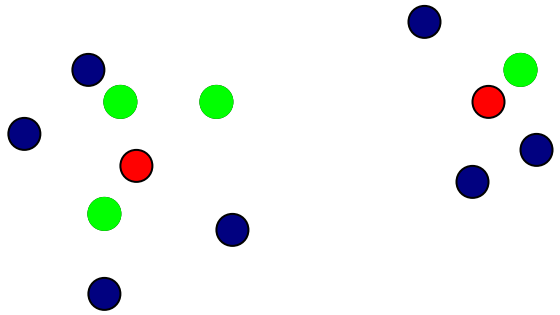
$$p(v) = \frac{1-\varepsilon}{n} + \varepsilon \sum_{u \in pr[v]} \frac{p(v)}{|su[u]|}$$

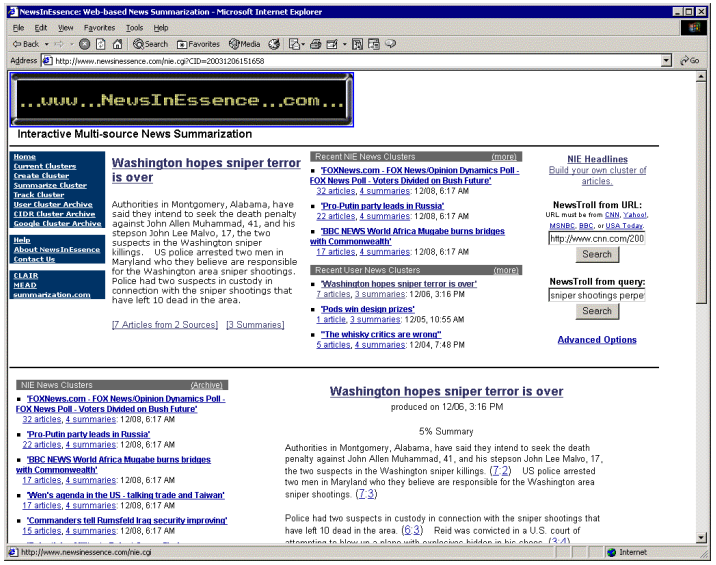
su = successor nodes
pr = predecessor nodes

- Eigenvector centrality: the paths in the random walk are weighted by the centrality of the nodes that the path connects

# The MEAD summarizer

- MEAD: salience-based extractive summarization (in 6 languages)
- Centroid-based summarization (single and multi document)
- Vector space model
- Additional features: position, length, lexrank
- Cross-document structure theory
- Reranker – similar to MMR
- 1000+ downloads from www.summarization.com/mead
- NIE: first robust news summarization system [Radev & al. 2001]





# Centrality in summarization

- Motivation: capture the most central words in a document or cluster
- Sentence salience [Boguraev & Kennedy 1999]
- Centroid score [Radev & al. 2000, 2004a]
- Alternative methods for computing centrality?

# LexPageRank (Cosine centrality)

## Example (cluster d1003t)

1 (d1s1) Iraqi Vice President Taha Yassin Ramadan announced today, Sunday, that Iraq refuses to back down from its decision to stop cooperating with disarmament inspectors before its demands are met.

2 (d2s1) Iraqi Vice president Taha Yassin Ramadan announced today, Thursday, that Iraq rejects cooperating with the United Nations except on the issue of lifting the blockade imposed upon it since the year 1990.

3 (d2s2) Ramadan told reporters in Baghdad that "Iraq cannot deal positively with whoever represents the Security Council unless there was a clear stance on the issue of lifting the blockade off of it.

4 (d2s3) Baghdad had decided late last October to completely cease cooperating with the inspectors of the United Nations Special Commission (UNSCOM), in charge of disarming Iraq's weapons, and whose work became very limited since the fifth of August, and announced it will not resume its cooperation with the Commission even if it were subjected to a military operation.

5 (d3s1) The Russian Foreign Minister, Igor Ivanov, warned today, Wednesday against using force against Iraq, which will destroy, according to him, seven years of difficult diplomatic work and will complicate the regional situation in the area.

6 (d3s2) Ivanov contended that carrying out air strikes against Iraq, who refuses to cooperate with the United Nations inspectors, ``will end the tremendous work achieved by the international group during the past seven years and will complicate the situation in the region."

7 (d3s3) Nevertheless, Ivanov stressed that Baghdad must resume working with the Special Commission in charge of disarming the Iraqi weapons of mass destruction (UNSCOM).

8 (d4s1) The Special Representative of the United Nations Secretary-General in Baghdad, Prakash Shah, announced today, Wednesday, after meeting with the Iraqi Deputy Prime Minister Tariq Aziz, that Iraq refuses to back down from its decision to cut off cooperation with the disarmament inspectors.

9 (d5s1) British Prime Minister Tony Blair said today, Sunday, that the crisis between the international community and Iraq ``did not end" and that Britain is still ``ready, prepared, and able to strike Iraq."

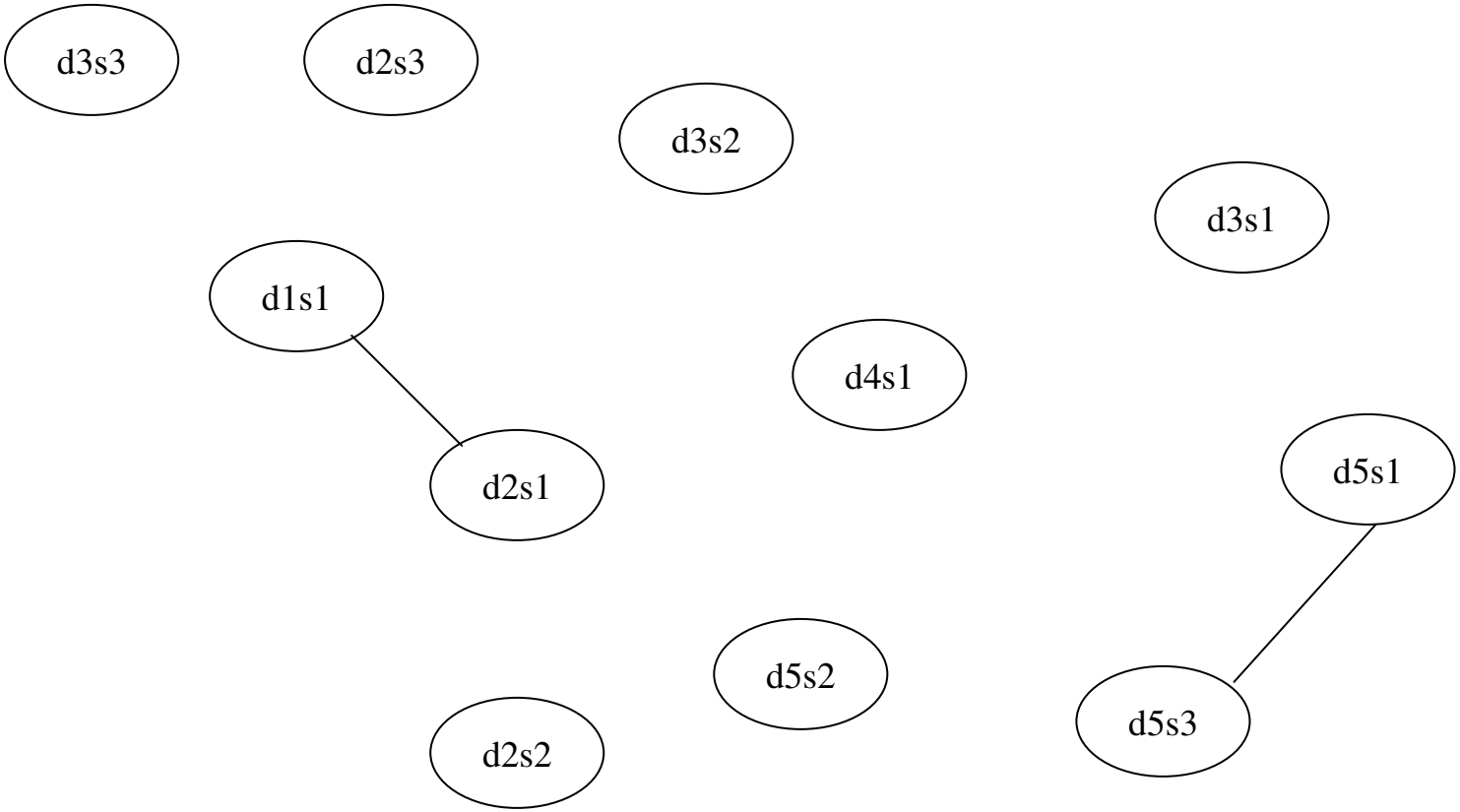
10 (d5s2) In a gathering with the press held at the Prime Minister's office, Blair contended that the crisis with Iraq ``will not end until Iraq has absolutely and unconditionally respected its commitments" towards the United Nations.

11 (d5s3) A spokesman for Tony Blair had indicated that the British Prime Minister gave permission to British Air Force Tornado planes stationed in Kuwait to join the aerial bombardment against Iraq.

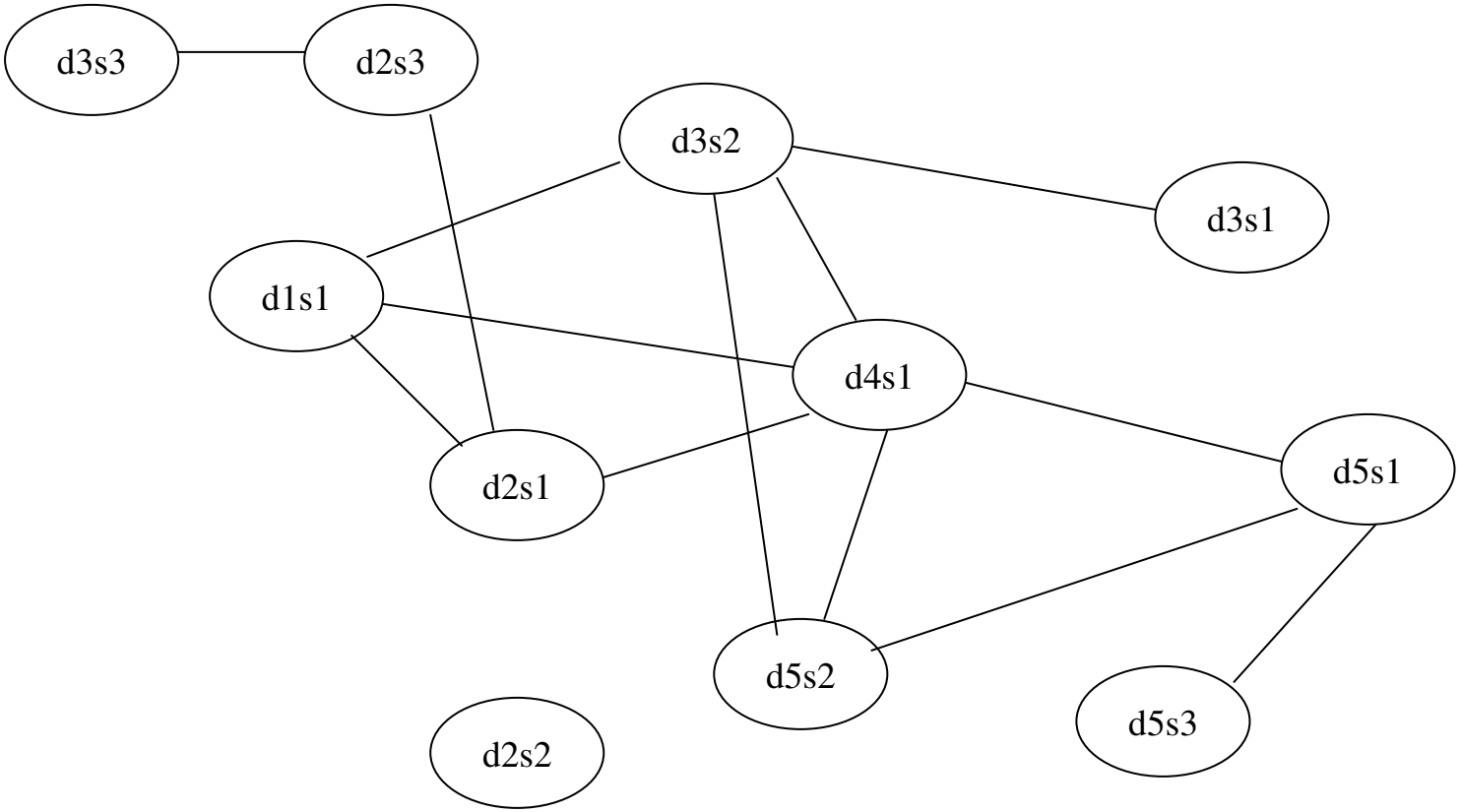
# Cosine centrality

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1.00 | 0.45 | 0.02 | 0.17 | 0.03 | 0.22 | 0.03 | 0.28 | 0.06 | 0.06 | 0.00 |
| 2 | 0.45 | 1.00 | 0.16 | 0.27 | 0.03 | 0.19 | 0.03 | 0.21 | 0.03 | 0.15 | 0.00 |
| 3 | 0.02 | 0.16 | 1.00 | 0.03 | 0.00 | 0.01 | 0.03 | 0.04 | 0.00 | 0.01 | 0.00 |
| 4 | 0.17 | 0.27 | 0.03 | 1.00 | 0.01 | 0.16 | 0.28 | 0.17 | 0.00 | 0.09 | 0.01 |
| 5 | 0.03 | 0.03 | 0.00 | 0.01 | 1.00 | 0.29 | 0.05 | 0.15 | 0.20 | 0.04 | 0.18 |
| 6 | 0.22 | 0.19 | 0.01 | 0.16 | 0.29 | 1.00 | 0.05 | 0.29 | 0.04 | 0.20 | 0.03 |
| 7 | 0.03 | 0.03 | 0.03 | 0.28 | 0.05 | 0.05 | 1.00 | 0.06 | 0.00 | 0.00 | 0.01 |
| 8 | 0.28 | 0.21 | 0.04 | 0.17 | 0.15 | 0.29 | 0.06 | 1.00 | 0.25 | 0.20 | 0.17 |
| 9 | 0.06 | 0.03 | 0.00 | 0.00 | 0.20 | 0.04 | 0.00 | 0.25 | 1.00 | 0.26 | 0.38 |
| 10 | 0.06 | 0.15 | 0.01 | 0.09 | 0.04 | 0.20 | 0.00 | 0.20 | 0.26 | 1.00 | 0.12 |
| 11 | 0.00 | 0.00 | 0.00 | 0.01 | 0.18 | 0.03 | 0.01 | 0.17 | 0.38 | 0.12 | 1.00 |

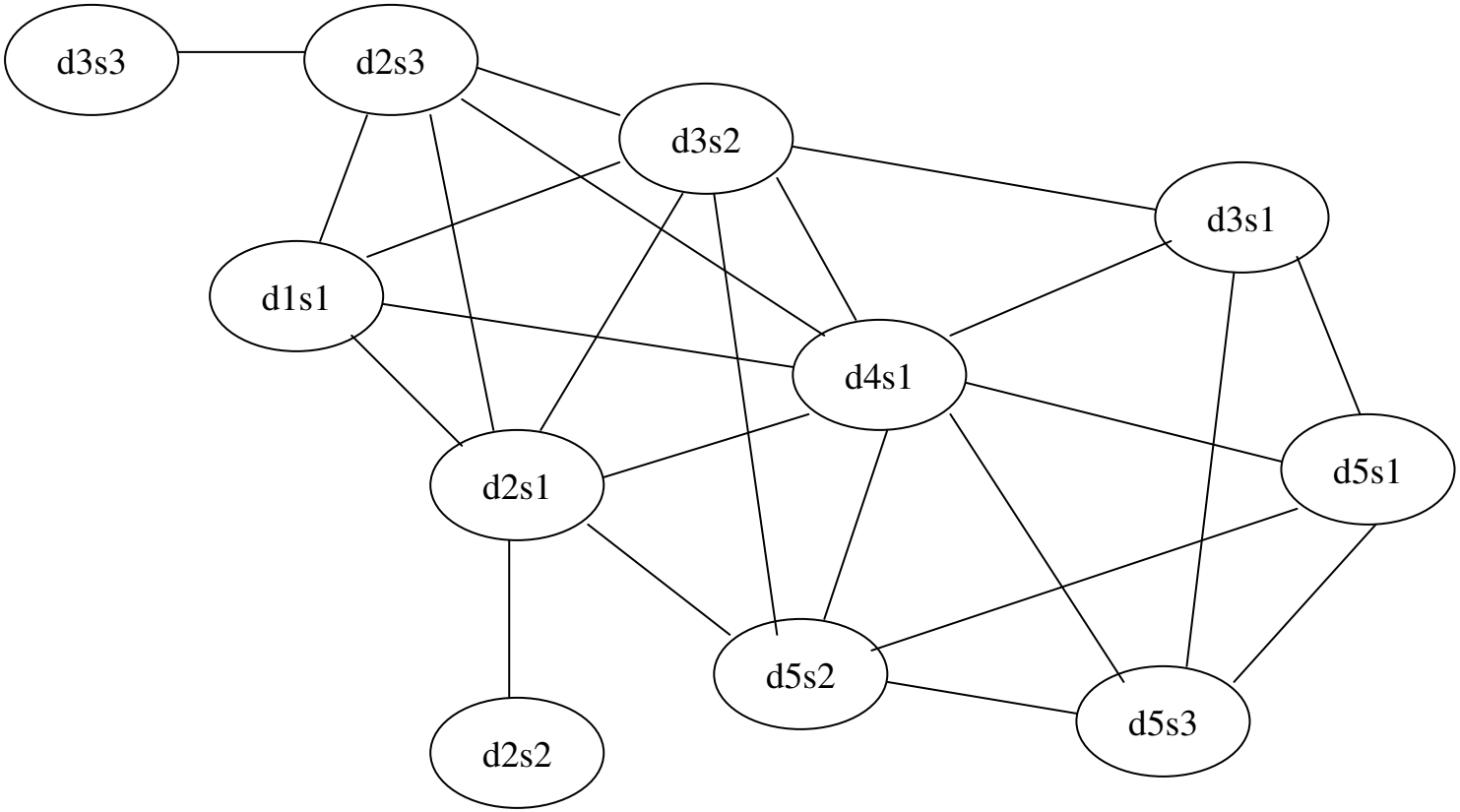
# Cosine centrality (t=0.3)



# Cosine centrality (t=0.2)



# Cosine centrality (t=0.1)




Sentences vote for the most central sentence!

# Cosine centrality vs. centroid centrality

| ID | LPR (0.1) | LPR (0.2) | LPR (0.3) | Centroid |
|---|---|---|---|---|
| d1s1 | 0.6007 | 0.6944 | 0.0909 | 0.7209 |
| d2s1 | 0.8466 | 0.7317 | 0.0909 | 0.7249 |
| d2s2 | 0.3491 | 0.6773 | 0.0909 | 0.1356 |
| d2s3 | 0.7520 | 0.6550 | 0.0909 | 0.5694 |
| d3s1 | 0.5907 | 0.4344 | 0.0909 | 0.6331 |
| d3s2 | 0.7993 | 0.8718 | 0.0909 | 0.7972 |
| d3s3 | 0.3548 | 0.4993 | 0.0909 | 0.3328 |
| d4s1 | 1.0000 | 1.0000 | 0.0909 | 0.9414 |
| d5s1 | 0.5921 | 0.7399 | 0.0909 | 0.9580 |
| d5s2 | 0.6910 | 0.6967 | 0.0909 | 1.0000 |
| d5s3 | 0.5921 | 0.4501 | 0.0909 | 0.7902 |

| | CODE | ROUGE-1 | ROUGE-2 | ROUGE-W |
|---|---|---|---|---|
| Centroid | C0.5 | 0.39013 | 0.10459 | 0.12202 |
| | C10 | 0.38539 | 0.10125 | 0.11870 |
| | C1.5 | 0.38074 | 0.09922 | 0.11804 |
| | C1 | 0.38181 | 0.10023 | 0.11909 |
| | C2.5 | 0.37985 | 0.10154 | 0.11917 |
| | C2 | 0.38001 | 0.09901 | 0.11772 |
| Degree | Degree0.5T0.1 | 0.39016 | 0.10831 | 0.12292 |
| | Degree0.5T0.2 | 0.39076 | 0.11026 | 0.12236 |
| | Degree0.5T0.3 | 0.38568 | 0.10818 | 0.12088 |
| | Degree1.5T0.1 | 0.38634 | 0.10882 | 0.12136 |
| | Degree1.5T0.2 | 0.39395 | 0.11360 | 0.12329 |
| | Degree1.5T0.3 | 0.38553 | 0.10683 | 0.12064 |
| | Degree1T0.1 | 0.38882 | 0.10812 | 0.12286 |
| | Degree1T0.2 | 0.39241 | 0.11298 | 0.12277 |
| | Degree1T0.3 | 0.38412 | 0.10568 | 0.11961 |
| LexPageRank | Lpr0.5T0.1 | 0.39369 | 0.10665 | 0.12287 |
| | Lpr0.5T0.2 | 0.38899 | 0.10891 | 0.12200 |
| | Lpr0.5t0.3 | 0.38667 | 0.10255 | 0.12244 |
| | Lpr1.5t0.1 | 0.39997 | 0.11030 | 0.12427 |
| | Lpr1.5t0.2 | 0.39970 | 0.11508 | 0.12422 |
| | Lpr1.5t0.3 | 0.38251 | 0.10610 | 0.12039 |
| | Lpr1T0.1 | 0.39312 | 0.10730 | 0.12274 |
| | Lpr1T0.2 | 0.39614 | 0.11266 | 0.12350 |
| | Lpr1T0.3 | 0.38777 | 0.10586 | 0.12157 |

# Some comments

- Very high results:
  - task 3 (very short summary of automatic translations from Arabic)
  - task 4 (short summary of automatic translations from Arabic) in all recall oriented measures
- Punctuation problems (with LCS: ROUGE-L and ROUGE-W)
- Task 2 – lower results due to a bug

# Results

| Peer code | Task | ROUGE-1 | ROUGE-2 | ROUGE-3 | ROUGE-4 | ROUGE-L | ROUGE-W |
|---|---|---|---|---|---|---|---|
| 141 | 3 | 5 | 2 | 1 | 1 | 2 | 2 |
| 142 | 3 | 5 | 1 | 1 | 1 | 4 | 3 |
| 143 | 4 | 1 | 2 | 1 | 1 | 6 | 6 |
| 144 | 4 | 3 | 1 | 1 | 1 | 7 | 7 |
| 145 | 4 | 1 | 2 | 2 | 2 | 4 | 4 |

Recall

LCS