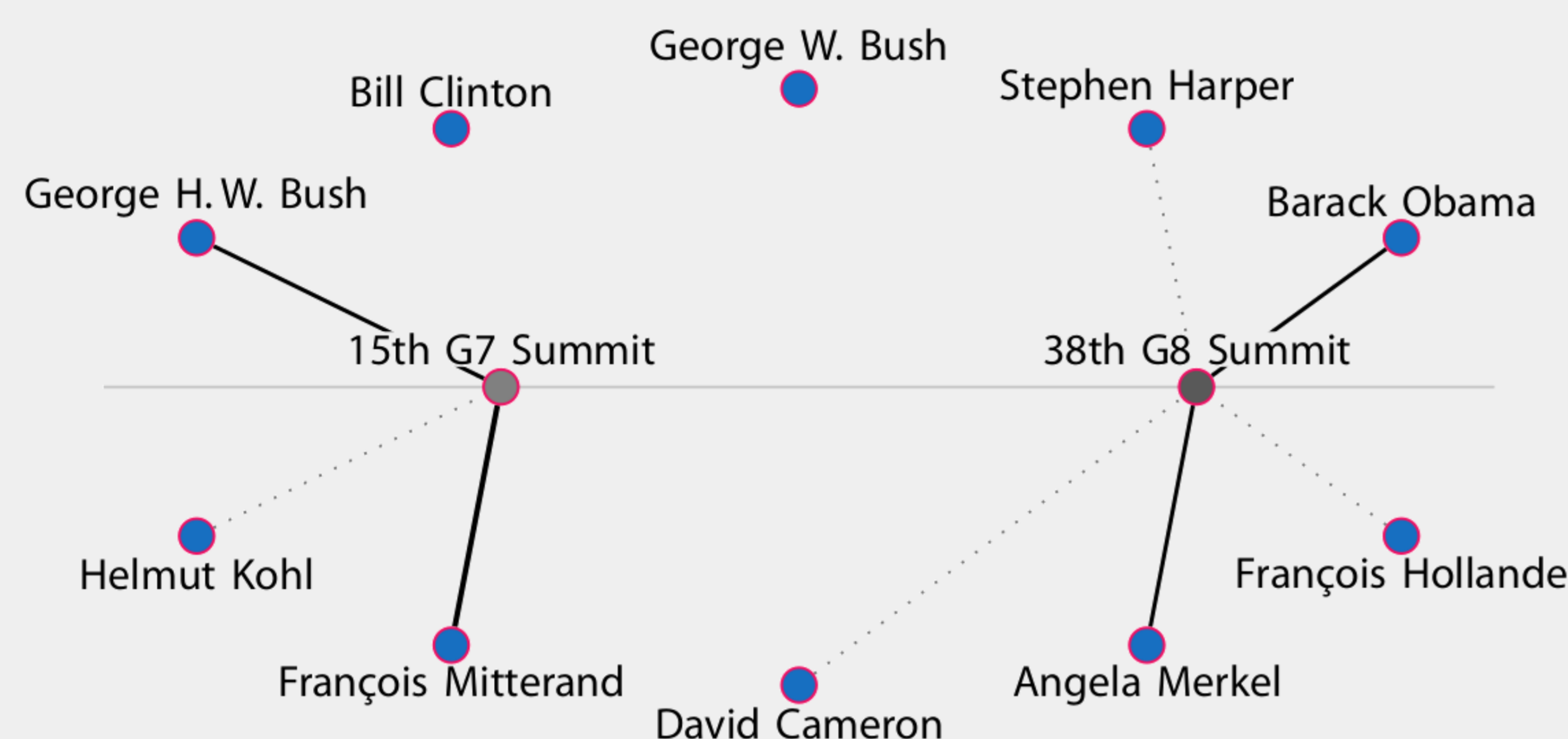


Efficient Computation of Relationship Centrality in Large Entity-Relationship Graphs

Stephan Seufert Srikanta Bedathur Johannes Hoffart
Andrey Gubichev Klaus Berberich

Relationship Centrality

Given a large Entity-Relationship graph (such as Wikipedia) and two sets of entities S, T , identify entities that play an important role in the relationship between S and T

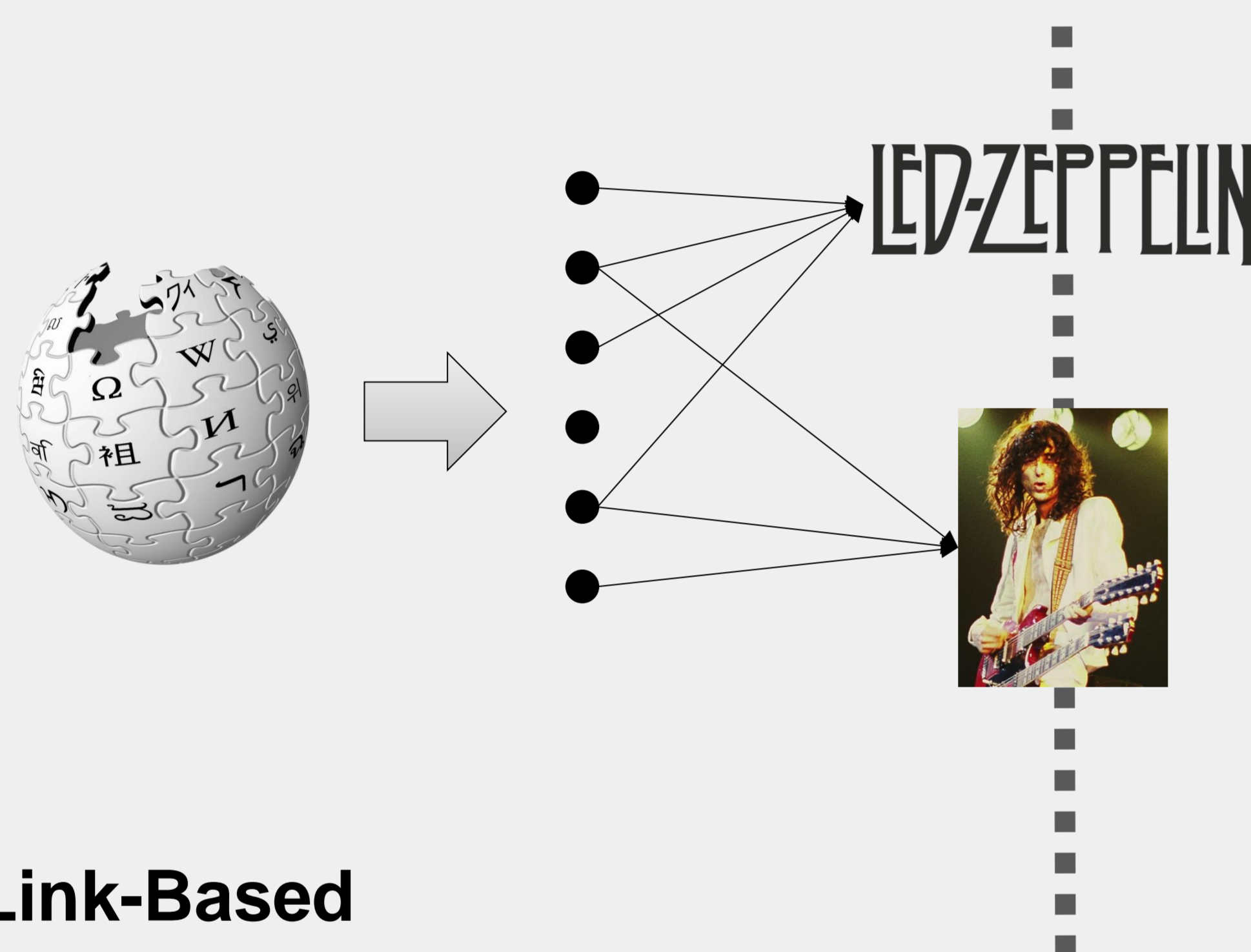


Ingredients:

- Underlying graph structure: Connectivity of the entities
- Semantic relatedness between individual entities to measure connection strength

Query sets can be specified manually or, when combined with a knowledge base such as YAGO, in the form of SPARQL queries.

Semantic Relatedness between Entities



	p	φ_e
γ_e	Blues Rock	0.2
	Atlantic Records	0.2
γ_f	Whole Lotta Love	0.2
	Rock guitarist	0.5

KORE: Keyphrase Overlap Relatedness



$$\text{KORE}(e, f) = \text{Blues Rock } \mathbf{0.2} \mid \mathbf{0.5} \text{ Rock guitarist}$$

$$\frac{\sum_{p \in P_e, q \in P_f} \text{PO}(p, q)^2 \cdot \min\{\varphi_e(p), \varphi_f(q)\}}{\sum_{p \in P_e} \varphi_e(p) + \sum_{q \in P_f} \varphi_f(q)}$$

$$\text{Rock } \mathbf{0.2} \mid \mathbf{0.3} \text{ Rock}$$

$$\text{PO}(p, q) = \frac{\sum_{w \in p \cap q} \min\{\gamma_e(w), \gamma_f(w)\}}{\sum_{w \in p \cup q} \max\{\gamma_e(w), \gamma_f(w)\}}$$

Link-Based

Count overlapping links
MW [Milne & Witten, WIKIAI 2008]

Keyphrase-Based

Intersect weighted keyphrases
KORE [Hoffart et al., CIKM 2012]

Measure

Centrality of vertex v with respect to the query sets S and T :

$$c_R(v) = \sum_{s \in S} \sum_{t \in T} \frac{1}{\rho(s, v, t)}$$

Using a connection *penalty* of the form

$$\rho(s, v, t) = (1 + d(s, v))(1 + d(v, t))$$

- Distance $d(s, v)$ captures the dissimilarity between the entities and depends on the semantic relatedness measure, for example

$$d(s, v) = 1 - \text{KORE}(s, v)$$

- Computation: $|S|+|T|$ rounds of Dijkstra's algorithm
- Speedup: Stop expansion after specified distance bound Δ .

Example Results

- Query 1: Events between European politicians and politicians from the United States
- Query 2: Movies between US action movie stars and Asian action movie stars
- Query 3: Events between countries from Middle East/Central Asia and Western countries

Rank	Entity
1	2009 G-20 Pittsburgh sum.
2	2010 G-20 Toronto summit
3	37th G8 summit
4	Iraq War
5	35th G8 summit
6	2009 G-20 London Summit
7	36th G8 summit
8	2010 G-20 Seoul summit
9	Presidency of G. W. Bush
10	2009 Nobel Peace Prize

Rank	Entity
1	The Expendables (2010 film)
2	Crouching Tiger, Hidden Dragon
3	The Forbidden Kingdom
4	Rush Hour 2
5	Police Story (1985 film)
6	Once Upon a Time in China II
7	Fist of Fury
8	Romeo Must Die
9	Kung Fu Hustle
10	Fearless (2006 film)

Rank	Entity
1	Iraq War
2	War in Afghanistan
3	Gulf War
4	Op. Enduring Freedom
5	Yom Kippur War
6	War on Terror
7	Battle of Karameh
8	Palestinian diaspora
9	Operation Opera
10	Suez Crisis

Efficiency

Query	$\Delta = \infty$	$\Delta = 1$	$\Delta = 0.5$	
	Time	Time	τ	Time
Q1	41,960.40 ms	18,629.80 ms	1.0	4,616.05 ms
Q2	48,174.80 ms	15,002.70 ms	1.0	5,117.02 ms
Q3	71,162.50 ms	32,028.50 ms	1.0	7,858.39 ms

- Execution times over the Wikipedia graph (~37M edges)
- Different choice of distance bound Δ

