Discovering Chance Scenarios using Small-World KeyGraphs and Evolutionary Computation

Xavier Llorà[1], Naohiro Matsumura[3],
David E. Goldberg[1], Yukio Ohsawa[2],
Kei Ohnishi[1], & Antonio Gonzales[1]

[1]Illinois Genetic Algorithms Lab, National Center for Supercomputer Application,
University of Illinois at Urbana-Champaign

[2]Chance Discovery Consortium, University of Tsukuba & University of Tokyo

[3]Graduate School of Economics, Osaka University.

xllora@illigal.ge.uiuc.edu

Motivation

- Chance Discovery using visual maps (KeyGraph) *(Ohsawa, Benson, & Yachida, 1998)*
- Chance interpretation is done by user
- An appropriate degree of complexity is required
- How can we obtain KeyGraphs that
  - are complex enough to contain chances?
  - present clear clusters and relations among them?
- The KeyGraph building process is sensitive to the parameters used

Approach to Comprehensible KeyGraphs Creation

- Search among the KeyGraph parameter space
- Components:
  - Scoring metric that evaluates the comprehensiveness of a given KeyGraph
  - Search algorithm
- Scoring metric:
  - Human judgment
  - Small-worldliness
- Evolutionary computation as search algorithm

Approach

- Three methods:
  - KeyGraph with default values *(Ohsawa, Benson, & Yachida, 1998)*
  - KeyGraph evolved using interaction with the user
  - KeyGraph evolved using small-worldliness
- Three document types:
  - Synopsis (Cyrano de Bergerac)
  - Descriptive (Degas' Absinthe painting)
  - Creative (Marketing scenario discussion)

Some Details on KeyGraph Building

- Document Processing (D')
  - Document compactation (stop-word removal and word stemming)
  - Phrase construction (most frequent work combinations)
- Extraction of high-frequency terms ($N_{hf} \subset D'$)
- Extracting links for all $N_{hf} \subset D'$ (top ranking association)

$$assoc(w_i, w_j) = \sum_{s \in D'} \min\left(|w_i|_s, |w_j|_s\right),$$

- Extracting key terms $K_{hk} \subset D'$ (connectors among clusters)

$$key(w) = 1 - \prod_{g \subset G}\left[1 - \frac{based(w,g)}{neighbors(g)}\right] \quad based(w,g) = \sum_{s \in D'} |w|_s, |g - w|_s,$$

$$neighbors(w) = \sum_{s \in D'} \sum_{w \in s} |w|_s, |g - w|_s,$$

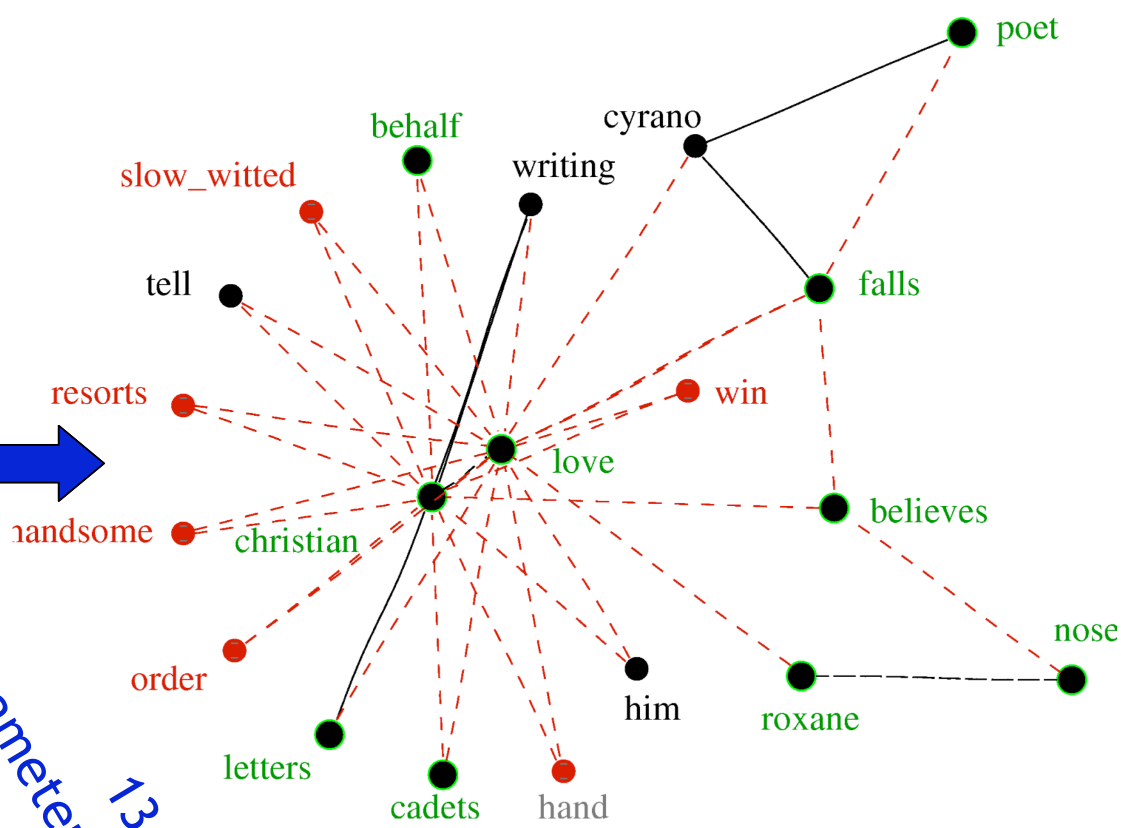
- Extracting key links among N_{hf} and N_{hf} using the *assoc* metric
- Keyword identify useful bridges among clusters ($N_{hf} \cup N_{hf}$)

A Simple KeyGraph Example

"A dashing officer of the guard and romantic poet, Cyrano de Bergerac falls in love with his cousin Roxane without her knowing. His one fault in his life, he feels, is his large nose and although it may have been a forming influence in his rapier-sharp wit, he believes that Roxane will reject him. He resorts to writing letters to her on behalf of one of his cadets, Christian, who is also in love with Roxane but just doesn't know how to tell her. She falls for the poetic charm of the letters but believes that they were written by Christian."

"The classic, tragic story brought to the screen once again. This French film tells the tale of the soulful poet philosopher and expert duelist named Cyrano who falls in love with the fair Roxane, but is ashamed to woo her because of his huge nose. Instead he writes love letters for slow-witted, but handsome Christian in order to win her hand for him. She falls deeply in love with the author, but doesn't know they were written by Cyrano."

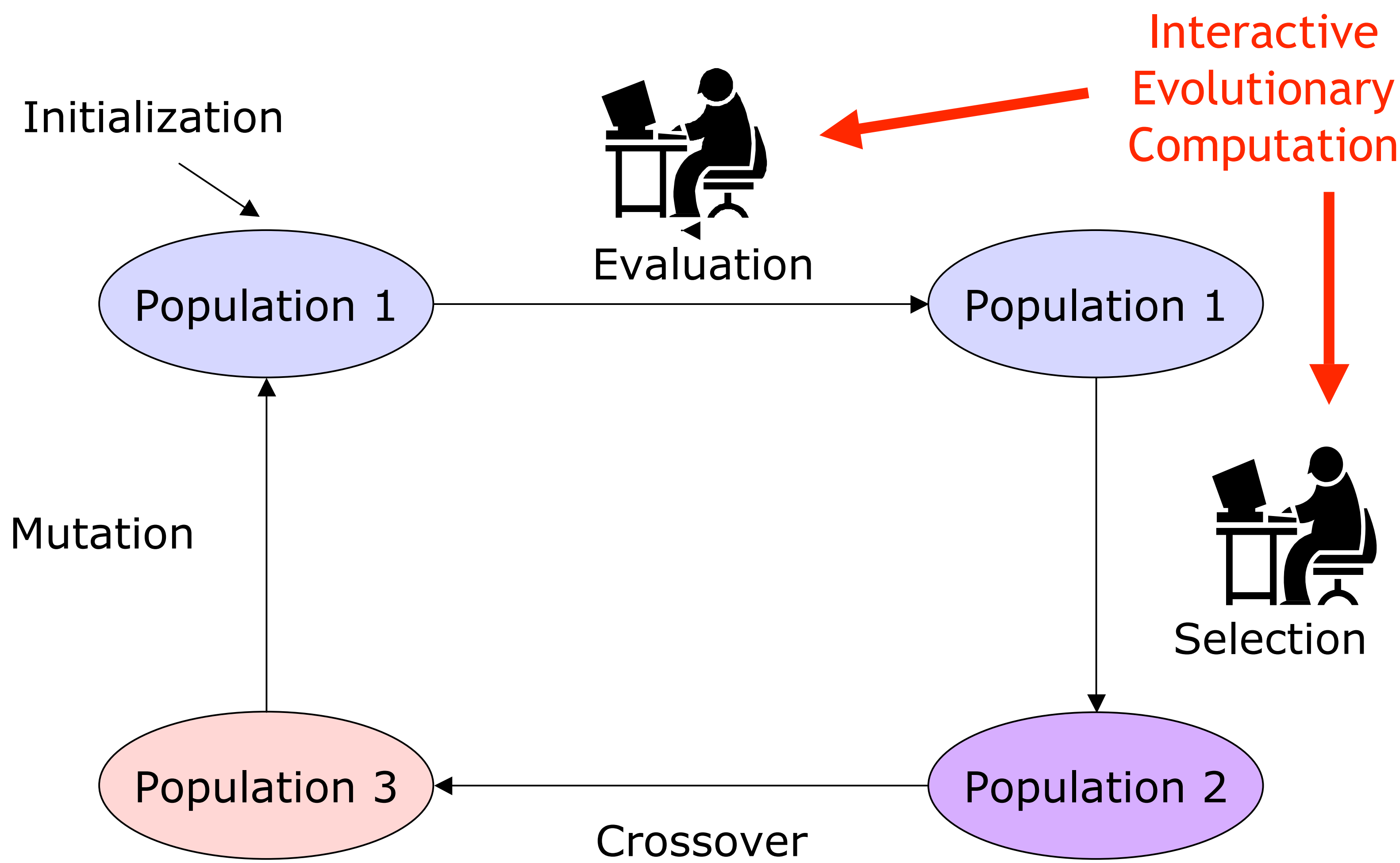
parameters 13



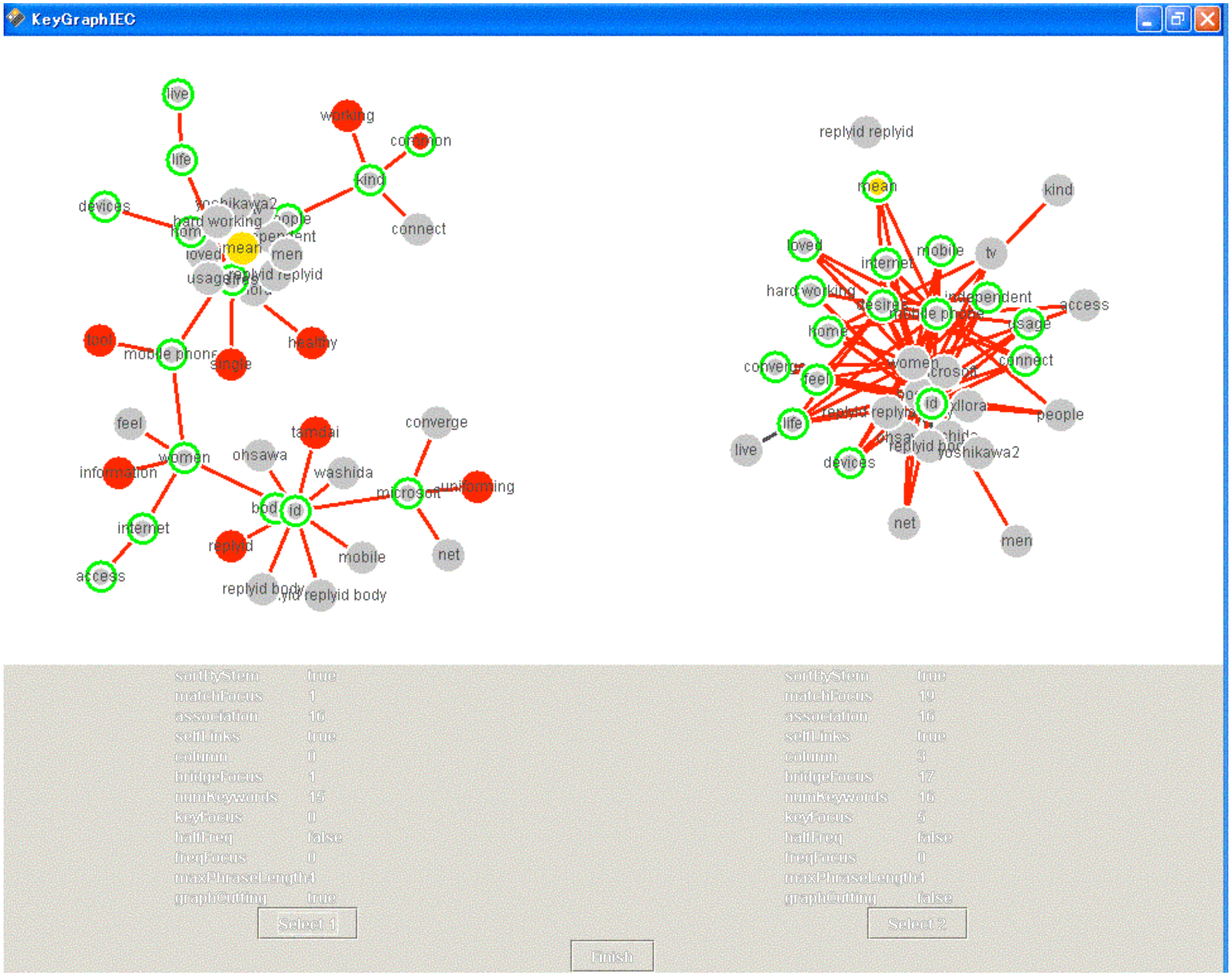
- High frequency terms
- Connect high frequency clusters
- Keywords

- Strongly correlated terms
- Significant bridges between clusters

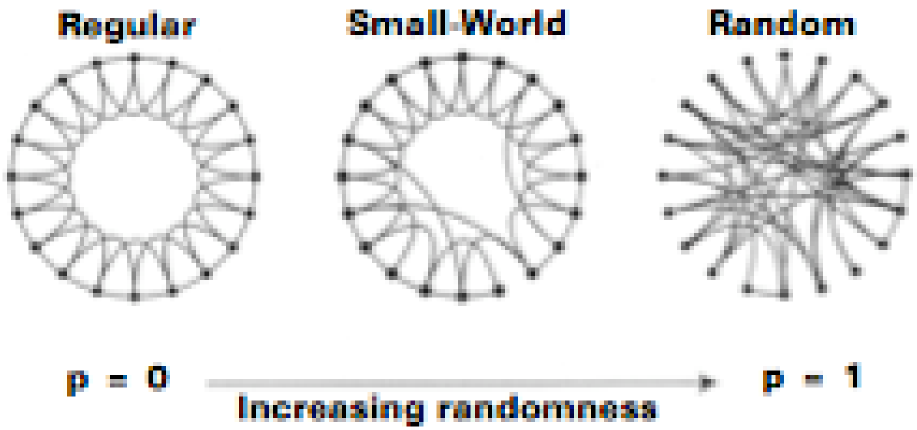
Evolutionary Computation



Interactive Evolutionary Computation



Small-world Topology



- A graph in which nodes are highly clustered yet the path length between them is small is called as small-world topology

Small-worldliness μ

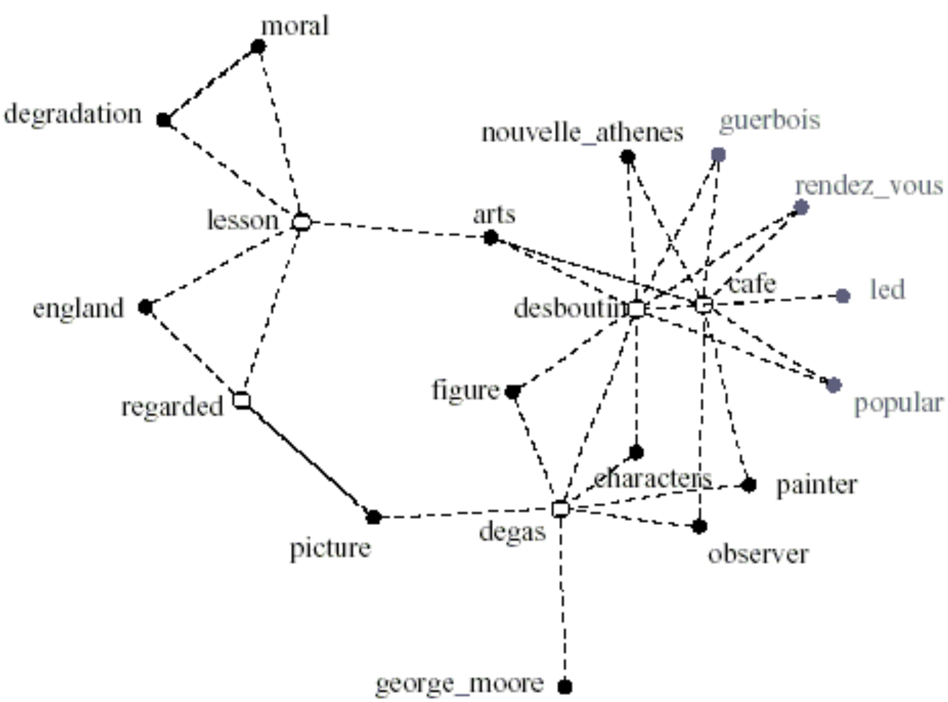
- The small-worldliness of a graph is obtained by **μ** ratio:

$$\mu = \frac{C / L}{C_{rand} / L_{rand}}$$

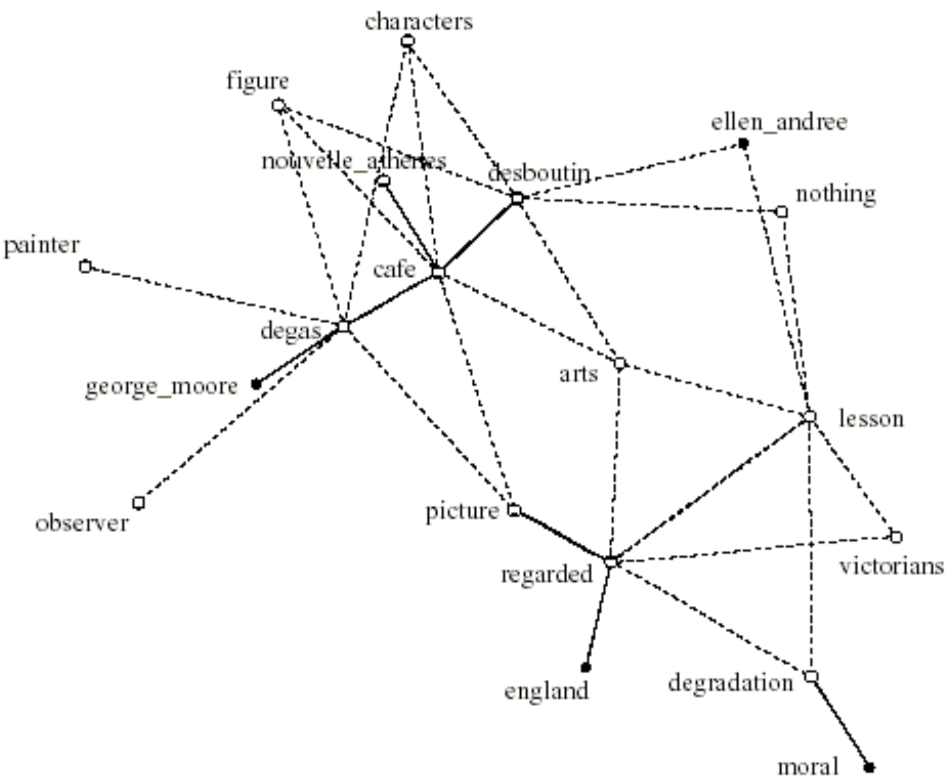
  where C is the clustering coefficient and L is the characteristic path length

- The ***μ*** ratio is used as the fitness function on a genetic algorithm

Results (i/iv)

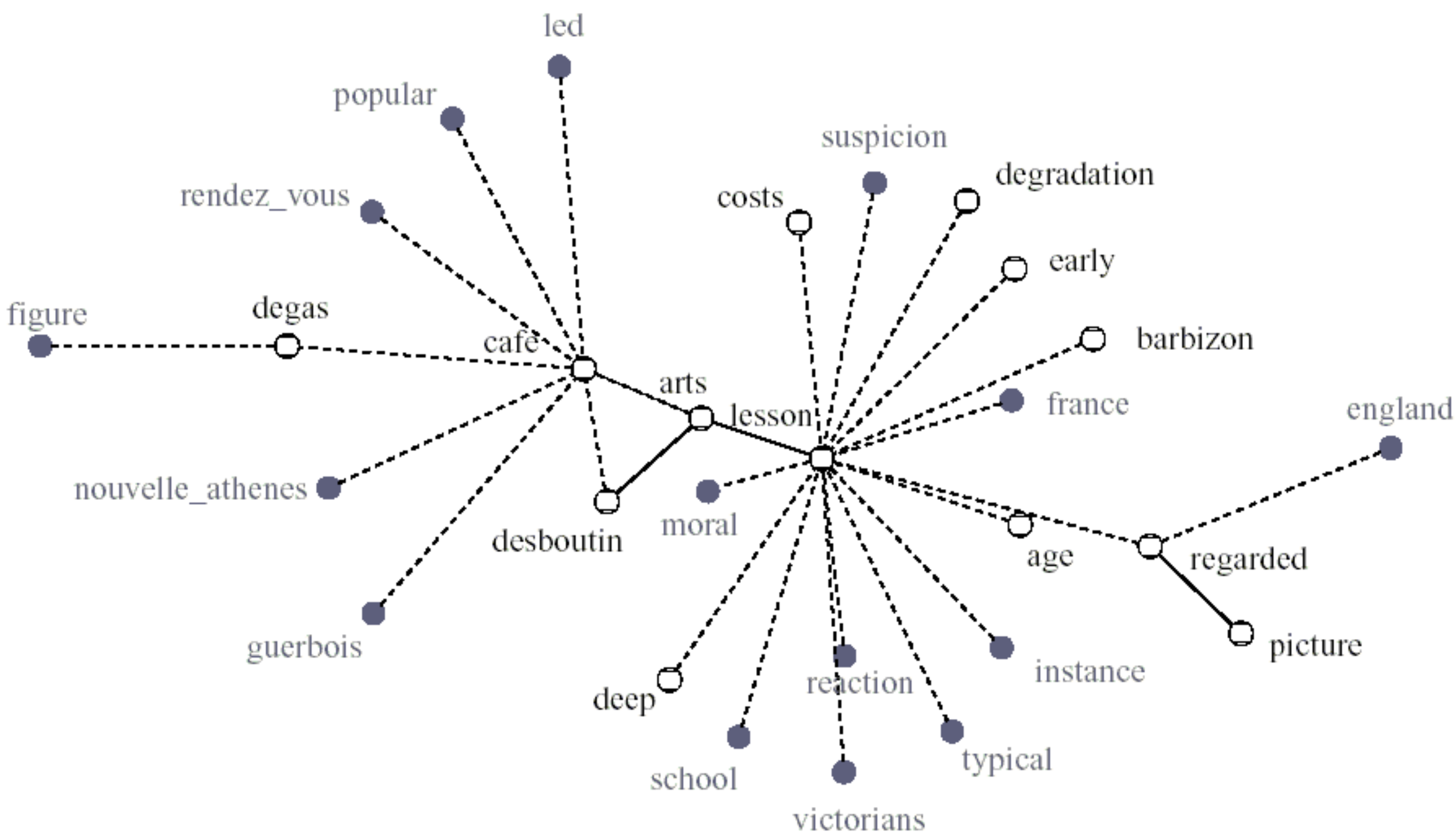


(c) History of Degas' Absinthe (default, $\mu = 1.07$)



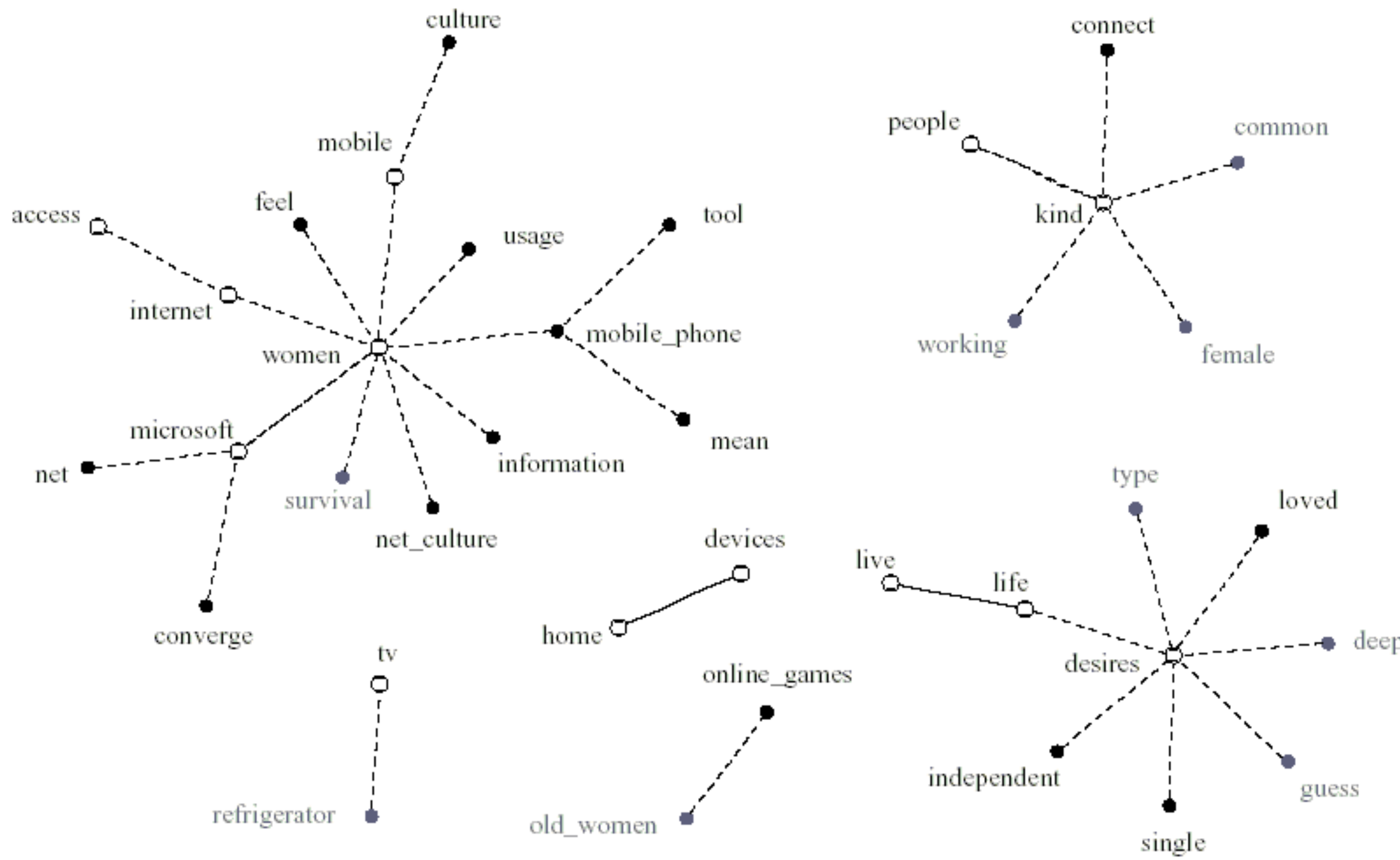
(d) History of Degas' Absinthe (iEC, $\mu = 6.18$)

Results (ii/iv)



(b) History of Degas' Absinthe ($\mu = 6.44$)

Results (iii/iv)



(b) KeyGraph tuned by means of an interactive evolutionary computation ($\mu = 1.83$)

Results (iv/iv)

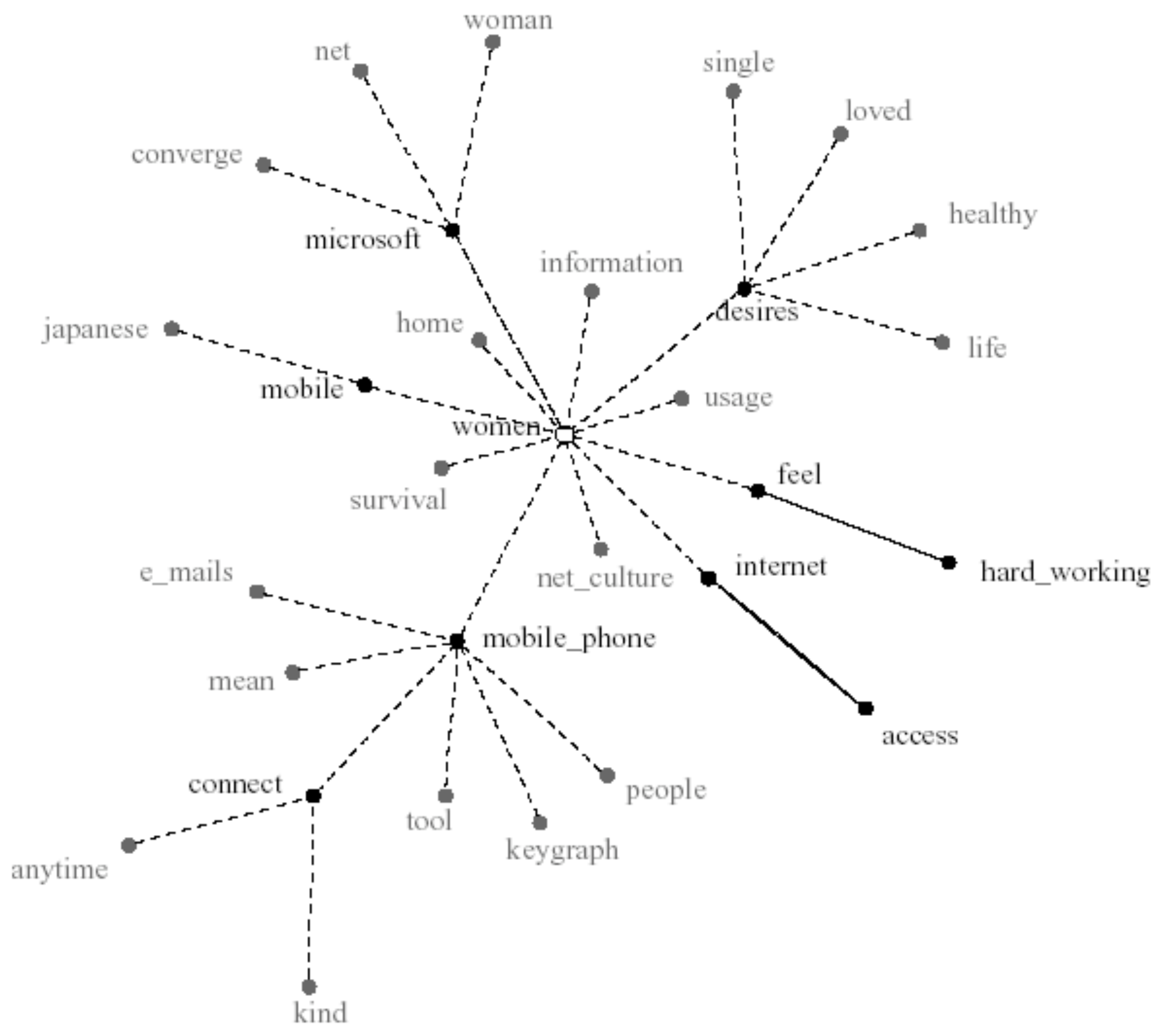


Figure 4: KeyGraph tuned using a genetic algorithm combined with a small world based fitness function ($\mu = 6.19$)

Discussion

Text	Method	C1	C2	C3	μ
Cyrano	KG	Difficult	Difficult	Difficult	4.07
Cyrano	KG+iEC	Medium	Medium	Medium	4.19
Cyrano	KG+SW+GA	Easy	Difficult	Difficult	10.7
Degas	KG	Medium	Medium	Medium	1.07
Degas	KG+iEC	Medium	Medium	Medium	6.18
Degas	KG+SW+GA	Easy	Easy	Easy	6.44
Marketing	KG	Difficult	Difficult	Difficult	2.07
Marketing	KG+iEC	Easy	Easy	Difficult	1.83
Marketing	KG+SW+GA	Easy	Easy	Easy	6.19

C1: ease to find clusters (except for the meanings)
C2: ease to understand the meaning of clusters
C3: ease to comprehend the relations among clusters

Conclusions

- First attempt to systematize the parameters tuning of KeyGraph
- Goal: create KeyGraphs where:
  - clusters are easy to find
  - clusters are easy to understand
  - the relations among clusters are easy to understand and help in the process of chance identification
- IEC provide and intuitive KeyGraph tuning
- Using small-worldliness provide intuitive graphs (requires a minimum amount of information)

Discovering Chance Scenarios using Small-World KeyGraphs and Evolutionary Computation

Xavier Llorà[1], Naohiro Matsumura[3],
David E. Goldberg[1], Yukio Ohsawa[2],
Kei Ohnishi[1], & Antonio Gonzales[1]

[1]Illinois Genetic Algorithms Lab, National Center for Supercomputer Application,
University of Illinois at Urbana-Champaign

[2]Chance Discovery Consortium, University of Tsukuba & University of Tokyo

[3]Graduate School of Economics, Osaka University.

xllora@illigal.ge.uiuc.edu