The Network Structure of Hard Combinatorial Landscapes

Marco Tomassini[1], Sebastien Verel[2], Gabriela Ochoa[3]

[1]University of Lausanne, Lausanne, Switzerland
[2]University of Nice Sophia-Antipolis , France
[3]University of Nottingham, Nottingham,UK

Motivation

- Use the tools of network analysis for studying:
  - The structure of combinatorial fitness landscapes
  - Problem (search) difficulty in combinatorial optimization
  - Design effective search algorithms
- How?
  - Mapping combinatorial landscapes to networks
  - Conduct a network analysis
  - Relate (and exploit?) network features to search operators

"the more we know of the statistical properties of a class of fitness landscapes, the better equipped will we be for the design of effective search algorithms for such landscapes"

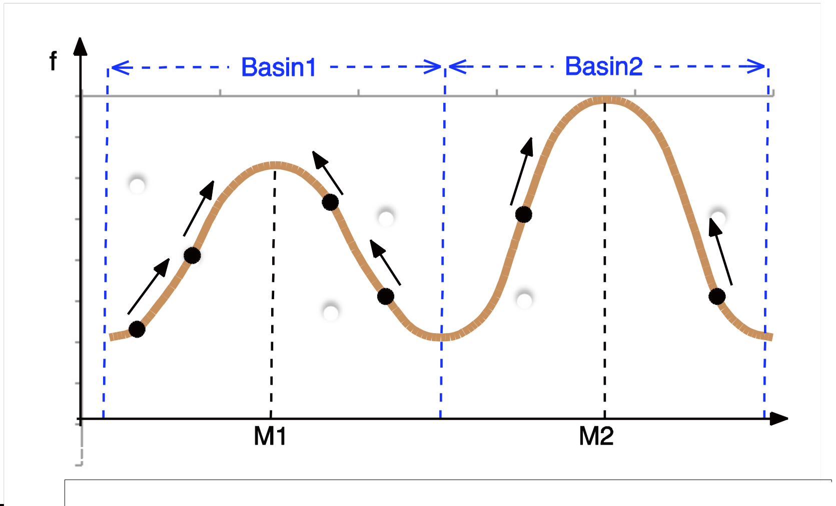
L. Barnett, U. Sussex, DPhil Diss. 2003

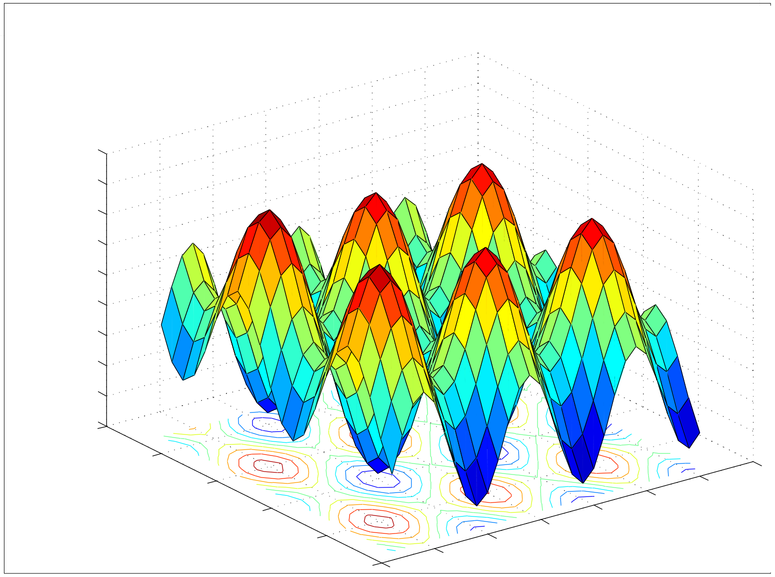
Outline

- Fitness landscapes
  - Features of fitness landscapes relevant to heuristic search
- Landscapes as networks
  - *Inherent networks* (energy surfaces)
  - Adapting this idea to combinatorial landscapes
    - How to define the nodes and the **edges**?
- Definitions and methods
- Results of the analysis
  - Basins of attraction
  - Network statistics and features

Fitness landscapes

- Describe dynamics of adaptation in Nature (Wright, 1932). Later, describe dynamics of evolutionary algorithms
- Search/Evolution: adaptive-walk over a Landscape
- 3 Components *L = (S,d,f)*
  - Search space
  - Neighborhood relation
  - Fitness function (height)

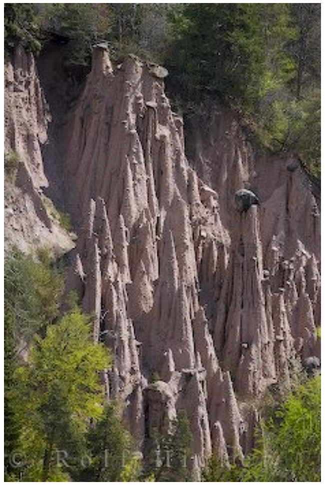




Features of landscapes relevant to heuristic search

- Number, fitness, distribution of local optima or peaks
- Topology of the basins of attraction
- Fitness differences between neighboring points (ruggedness)
- Presence and structure of *neutral networks* (terrains with equal fitness)

M. Fuji, Japan



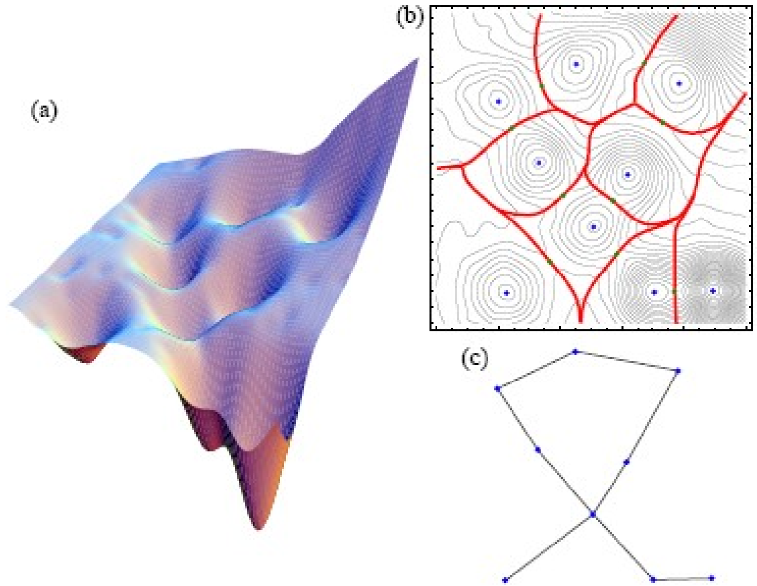
Earth pyramids, Tyrol, Italy

M. Auyantepui, Venezuela
(Angel Falls, Highest Waterfall)

Landscapes as networks (energy surfaces)

Inherent Network

- Vertices: energy minima
- Edges: two nodes are connected if the energy barrier separating them is sufficiently low (transition state)



(a) Model of 2D energy surface

(b) Contour plot, partition of the configuration space into basins of attraction surrounding minima

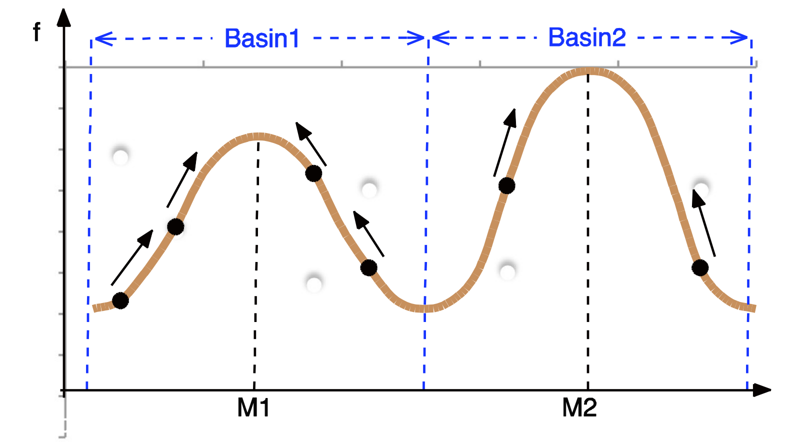
(c) landscape as a network

J. P. K. Doye, The network topology of a potential energy landscape: a static scale-free network, *Phys. Rev. Lett.*, 88, 2002

Landscapes as networks (combinatorial landscapes)

Definitions

- *Local optimum*: is a solution s^*, such that for all s in $V(s)$, $f(s)<f(s^*)$
- In practice: found by running a *best-improvement* local search (*LS*)
- *LS* defines a mapping from the search space S to the set of locally optimal solutions S^*
- Basin of attraction: b_i of a local optimum i is the set of configurations s in S, s.t. *LS* from s, will end in i

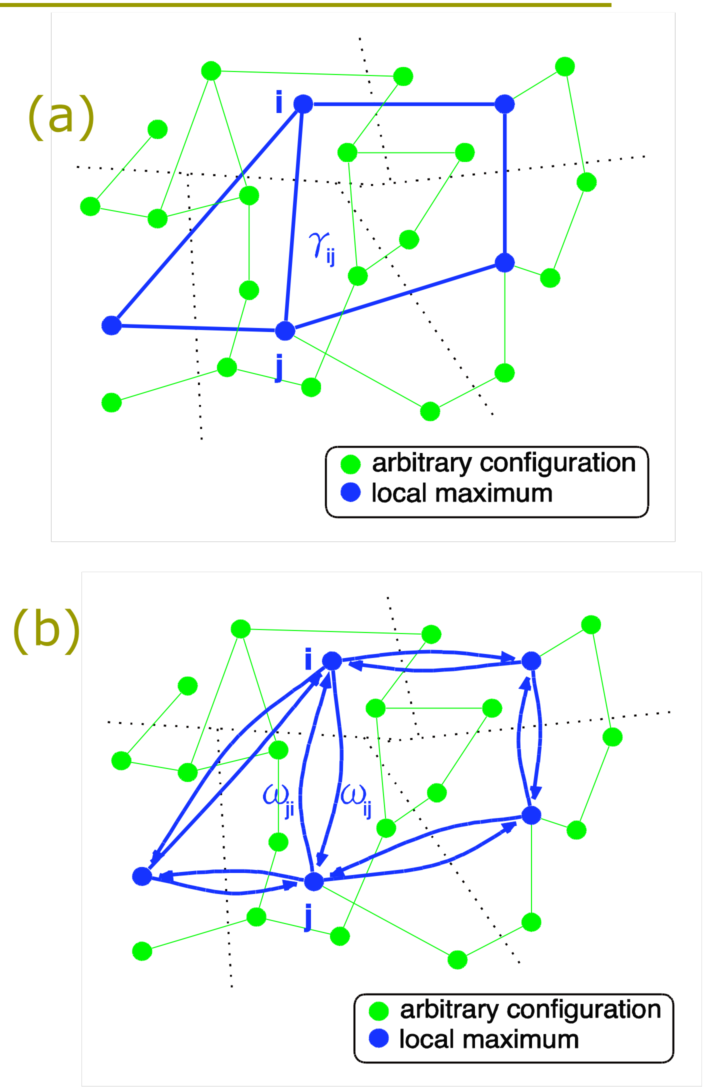


binary strings of length N. Neighbourhood $V(s)$ defined by the 1-move or bit flip operator $|V(s)| = N$

Landscapes as networks (combinatorial landscapes)

Local optima Network G = (S,E)*

- Nodes: S*
- Edges: notion of *connectivity* between basins
  - e_{ij} between *i* and *j*, if there is at least a pair of direct neighbours s_i and s_j, s.t. s_i in b_i, s_j in b_j (GECCO, 2008)
  - weights w_{ij} is attached to the edges, account for transition probabilities between basins (ALIFE, 2008, Phys. Rev. E,2008)
    - Two weights are needed, oriented transition graph



Definition - edge weight

- Partition of the configuration space S: $S = \cup_{i \in S*} b_i$ and $\forall i \in S \; \forall j \neq i$, $b_i \cap b_j = \emptyset$
- For each s and s', $p(s \to s')$ as the probability to pass from s to s'
- there are N neighbors for each solution, therefore:

  if $s' \in V(s)$, $p(s \to s') = \frac{1}{N}$ and
  if $s' \notin V(s)$, $p(s \to s') = 0$.
- Probability to pass from a solution $s \in S$ to a solution belonging to the basin b_j, as:

$$p(s \to b_j) = \sum_{s' \in b_j} p(s \to s')$$

- Total probability of going from basin b_i to basin b_j is the average over all $s \in b_i$ of the transition probabilities to solutions $s' \in b_j$:

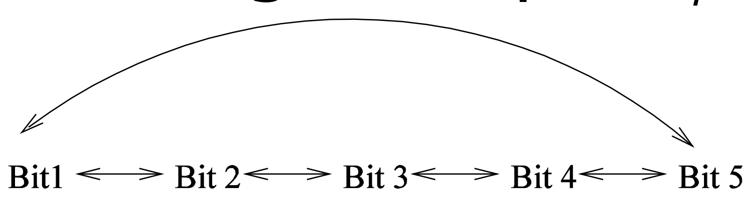
$$p(b_i \to b_j) = \frac{1}{\sharp b_i} \sum_{s \in b_i} p(s \to b_j)$$

NK landscapes (Kauffman, 93)

- A string s of N "spins" or "genes" represented by binary variables $s_i \in \{0, 1\}$
- A real stochastic function $\Phi : s \rightarrow \mathbb{R}_+$
- K ($0 \leq K < N$) determines how many other spins in the string influence a given spin s_i

The value of Φ is the average contribution of all the spins

$$\Phi(s) = \frac{1}{N} \sum_{i=1}^{N} \phi_i(s_i, s_{i_1}, \ldots, s_{i_K})$$



N=5, K = 2, Adjacent interaction

NK landscapes ctd.

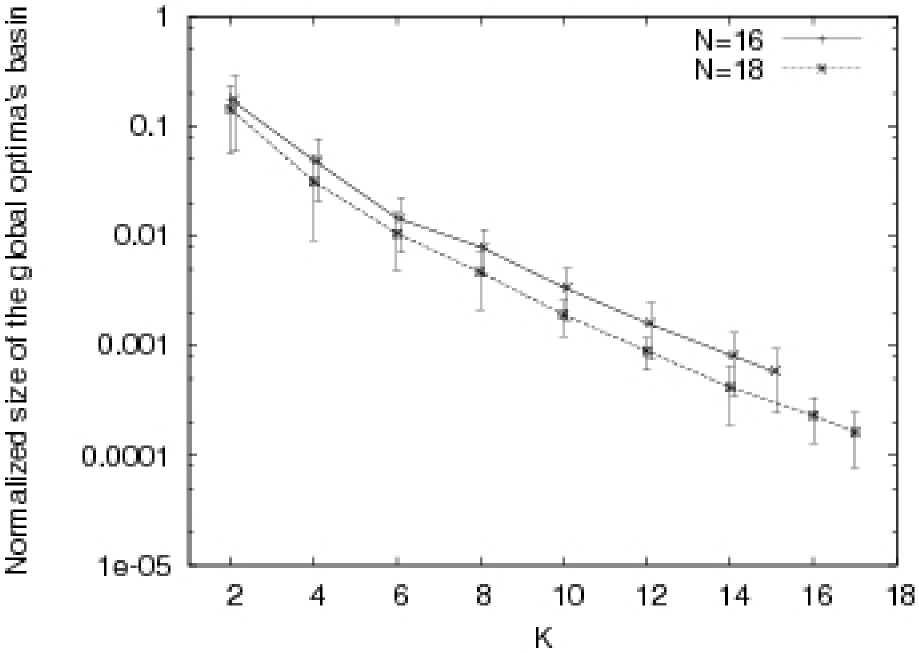
- K from 0 to $N - 1$, NK landscapes can be tuned from"smooth" to "rugged" (easy to difficult respectively)
- $K = 0$ no correlations, Φ is an additive function, and there is a single maximum
- $K = N - 1$ landscape completely random, the expected number of local optima is $2^N/(N + 1)$
- Intermediate values of K interpolate between these two extreme cases and have a variable degree of epistasis (i.e. gene interaction)

Methods

- Extracted and analysed networks for N=14, 16 and 18, K=2, 4, … ,N-2, N-1 (30 random instances for each case)
- We measured:
  - Statistics on basins sizes and fitness of optima
  - Network features: clustering coefficient, shortest path to the global optimum, weight distribution, disparity, boundary of basins

Analysis of basins
(Global optimum basin size versus K)

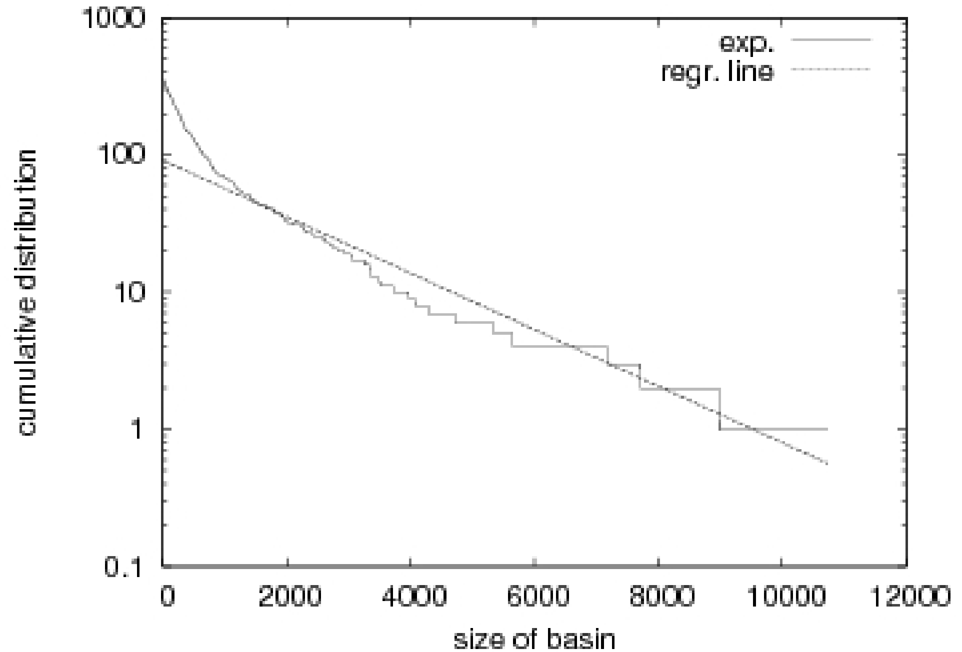
- Trend: the basin shrinks very quickly with increasing K.
- for higher K, more difficult for a search algorithm to locate the basin of attraction of the global optimum



Size of the basin corresponding to the global maximum for each K

Analysis of basins
basin size

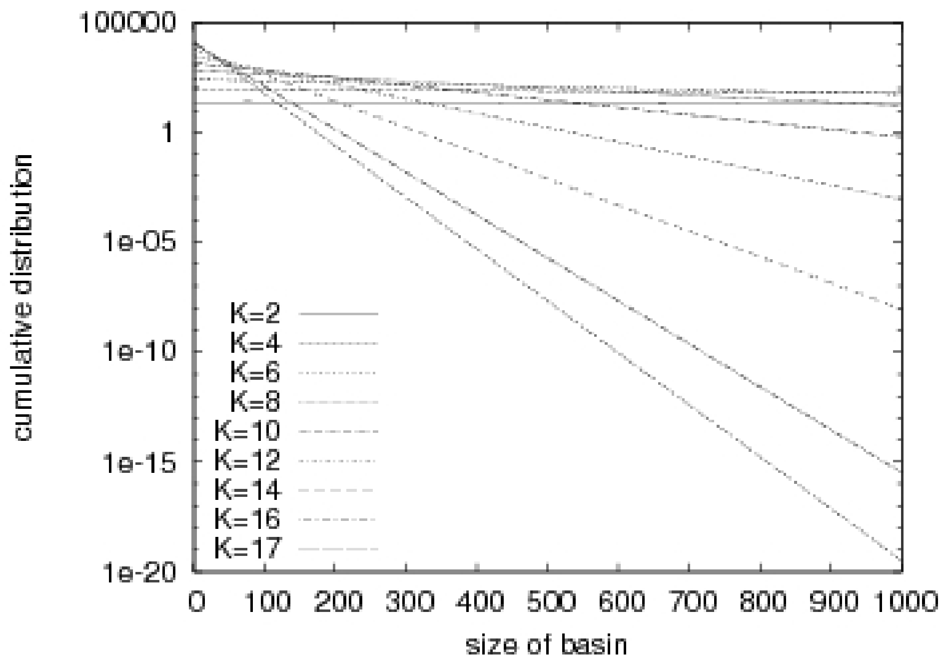
- Trend: small number of large basin, large number of small basin
- log-normal cumulative distribution
- slope of correlation increases with K
- when K large : basin sizes are nearly equals



Cumulative distribution of basins sizes for N=18 and K=4

Analysis of basins
basin size

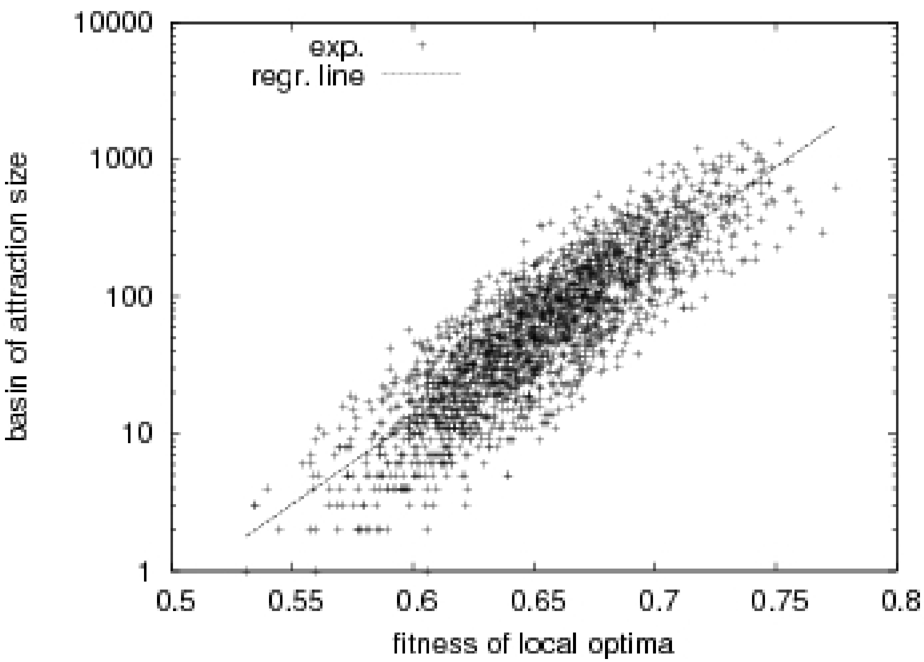
- Trend: small number of large basin, large number of small basin
- log-normal cumulative distribution
- slope of correlation increases with K
- when K large : basin sizes are nearly equals



Regression lines for N=18 and different values of K

Analysis of basins (fitness vs. basin size)

- Trend: clear positive correlation between the fitness values of maxima and their basins' sizes
- On average, the global optimum easier to find than one other local optimum
- But more difficult to find, as the number of local optima increases exponentially with increasing K



Correlation fitness of local optima vs. their corresponding basins sizes

General network statistics
(Example data for N = 16, avg. of 30 instances)

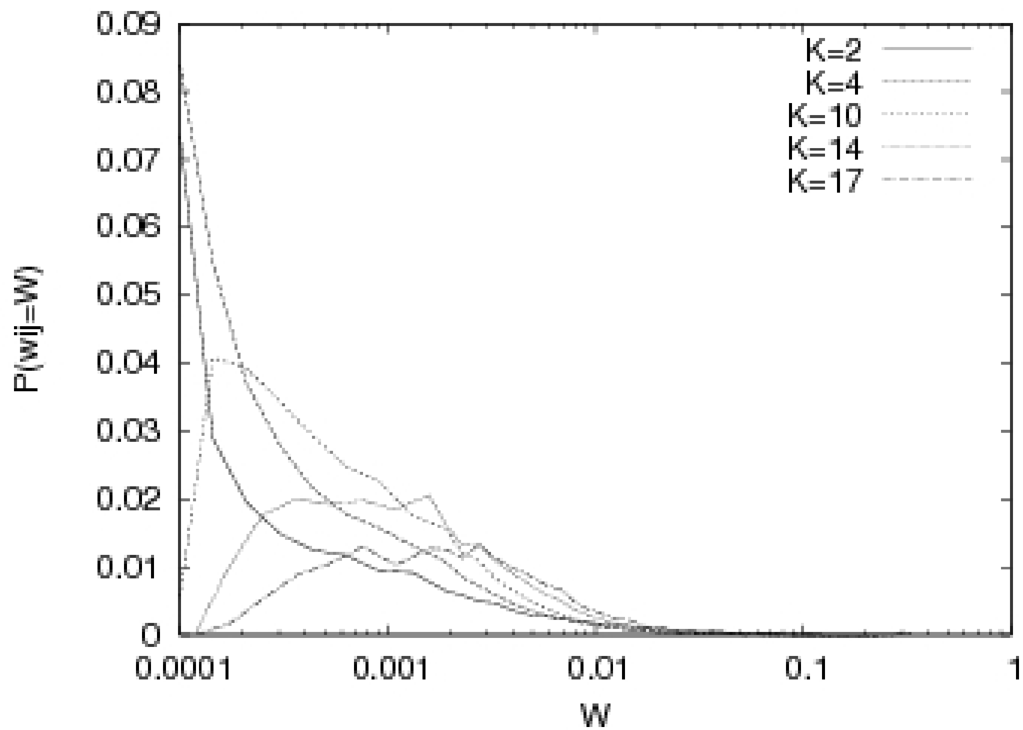
K	# nodes	# edges	clustering	disparity	path length
2	33	516	0.96	0.326	56
6	460	41,791	0.79	0.084	170
10	1,470	16,2139	0.53	0.050	206
15	3,868	32,1203	0.35	0.039	200

Clustering: For high K, transition between a given pair of neighboring basins is less likely to occur

Disparity (i.e. dishomogeneity of nodes with a given degree**)**: For high K the transitions to other basins tend to become equally likely, an indication of the randomness of the landscape

Weight distribution

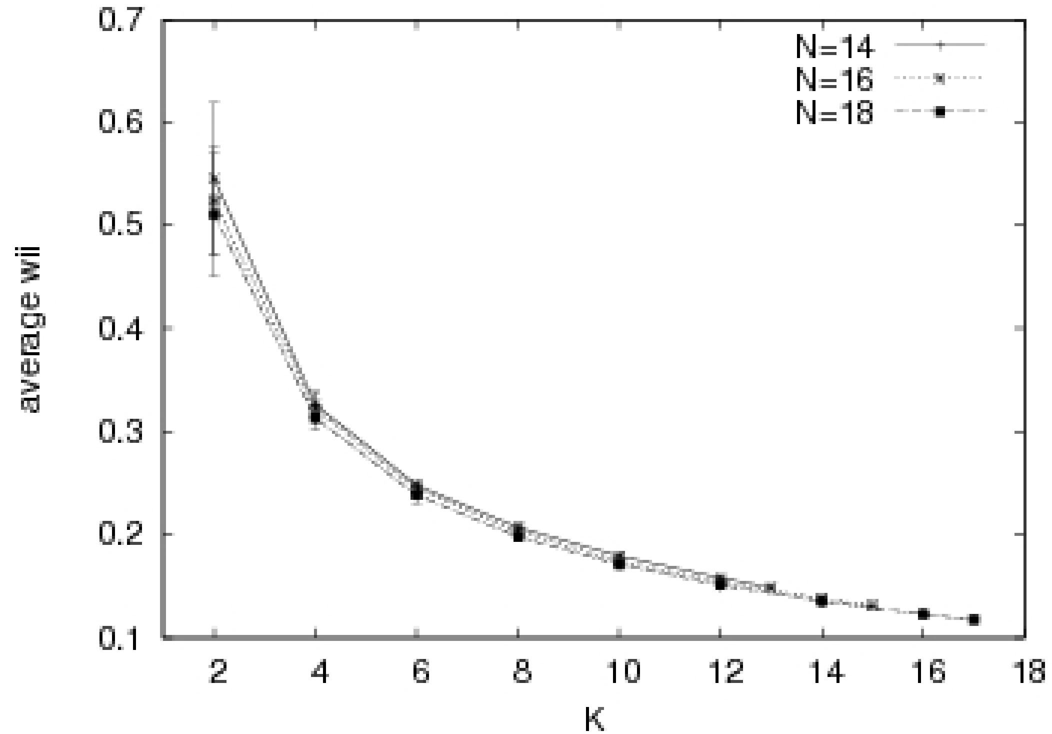
- Weights (transition prob. between neighbouring basins) are small
- For high K the decay is faster
- Low K has longer tails
- On average, the transition probabilities are higher for low K



distribution of the network weights *wij* for outgoing edges with $j \neq i$ in log-x scale, N = 18

Weight distribution remain in the same basin

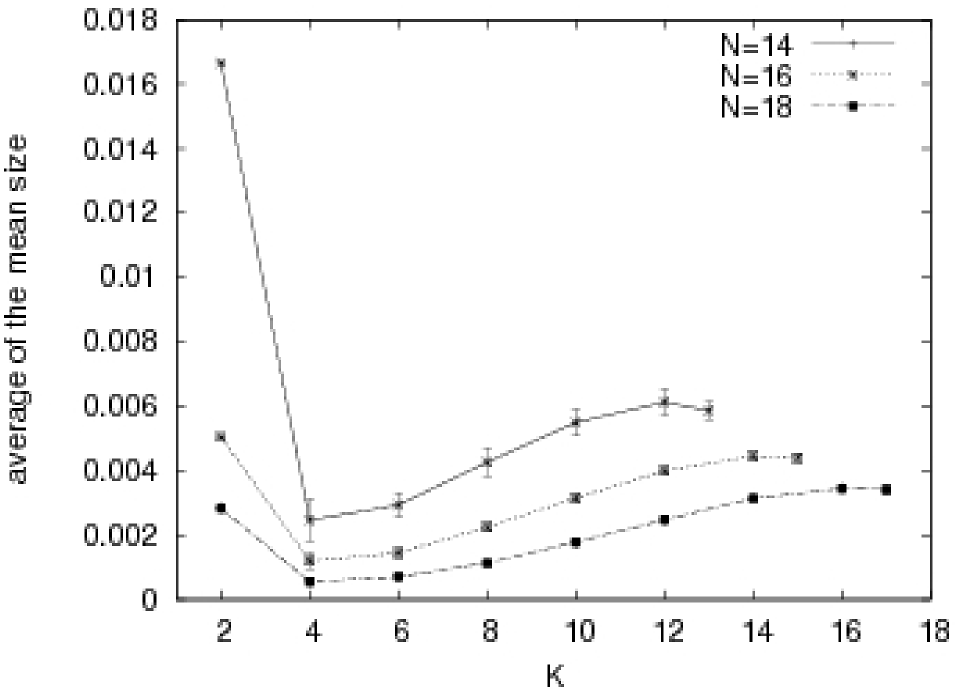
- Weights to remains in the same are large compare to Wij with i != j
- Wii are higher for low K
- Easier to leave the basin for high K : high exploration
- But : number of local optima increases fast with K



Average weight Wii according to the parameter N and K

Weight distribution
Size of basin interior

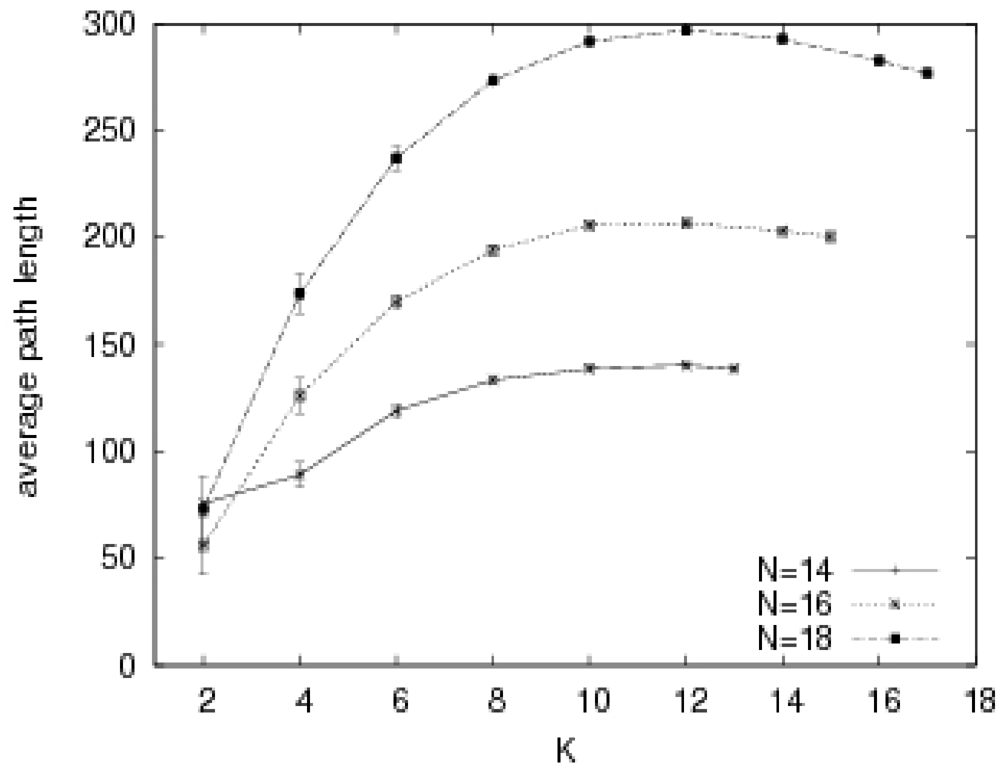
- Do the basins look like a “montain” with interior and border ?
- solution is in the interior if all neighbors are in the same basin
- Interior is very small
- Nearly all solution are in the border



Average of the mean size of basins interiors

Shortest path length between local optima

Average distance (shortest path) between nodes

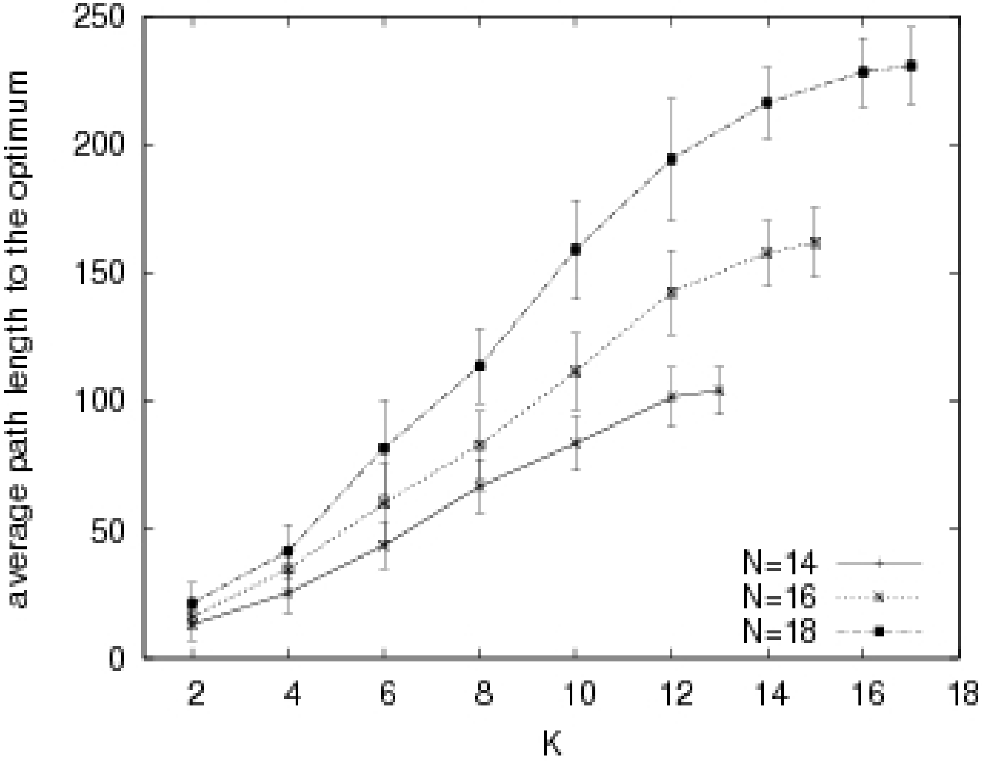


- Increase with N (# of nodes increase exponentially)
- For a given N, increase with K up to K = 10, then stagnates

Shortest path length to global optima

Average path length to the global optimum from all the other basins

- More relevant for optimisation
- Increase steadily with increasing K



Summary

- Proposed characterization of combinatorial landscapes as networks
- New findings about basin’s structure
- Related some network features to search difficulty
- Future?
  - Sampling techniques (instead of exhaustive enumeration)
  - Neutral and more realistic combinatorial landscapes

References

- M. Barthelemy, A. Barrat, R. Pastor-Satorras, and A. Vespignani, Characterization and modeling of weighted networks, *Physica A* 346 (2005), 34–43.
- J. P. K. Doye, The network topology of a potential energy landscape: a static scale-free network, *Phys. Rev. Lett*. 88 (2002), 238701.
- J. P. K. Doye and C. P. Massen, Characterizing the network topology of the energy landscapes of atomic clusters, J. *Chem. Phys.* 122 (2005), 084105.
- G. Ochoa, M. Tomassini, and S. Verel, A study of NK landscape sand basins and local optima networks, Genetic and Evolutionary Computation Conference, GECCO 2008, Proceedings, ACM, 2008
- S. Verel, G. Ochoa, M. Tomassini (2008) The Connectivity of NK Landscapes' Basins: A Network Analysis, Artificial Life XI, MIT Press, Cambridge, MA, pp. 648-655.
- M. Tomassini, S. Verel, G. Ochoa (2008) Complex-network analysis of combinatorial spaces: the landscape case of NK landscapes. Physical Review Letter E, 78 (6), 066114, 2008.