Laboratoire I3S, CNRS/Université de Nice-Sophia Antipolis

From Royal Road To Epistatic Road For Variable Length Evolution Algorithm

Michael Defoin Platel **Sébastien Verel**

Manuel Clergue **Philippe Collard**

June 2003

1 – Introduction

- In GA, fixed Lenght Representation is used : Each position of the genotypes corresponds to one gene
- In messy GA or linear GP, Variable Lenght Representation (VLR) is used : Genotypes have a variable number of genes identified by their forms

Some specific obstacles come with the variable length paradigm :

- Exploration of sizes
- Identification of genes
- Neutrality (position of genes, intron)

1 – Introduction

One way to understand difficulty in EA : Design a problem with parameters controlling main features of landscape (Royal Road, NK-landscapes...)
After, these problem can used as benchmark

With VLR, there are only a few attempts to design academic frameworks (the Royal Tree and the Royal Road for Linear GP)

We propose Epistatic Road (ER) :
inspired from Royal Road (for VLR) and NK-landscapes

2 – Royal Road for variable length representation

Study the destructive (or constructive) effects of crossover on building blocks in VLR (Defoin 2003)

2.1 – Definition

Set of optima : $\{g \in G_\Sigma \mid \forall l \in \Sigma,\ B_b(g,l) = 1\}$ with :

$$B_b(g,l) = \begin{cases} 1 & \text{if } \exists i \in [0, \lambda - b] \mid \forall j \in [0, b-1],\ g_{i+j} = l \\ 0 & \text{else} \end{cases}$$

- $b \geq 1$ the size of blocks
- Σ an alphabet of size N that defines the set of all possible letters l per locus
- G_Σ the set of all genotypes of size $\lambda \leq \lambda_{max}$ defined over Σ
- g a genotype of size $\lambda \leq \lambda_{max}$
- g_k the k^{th} locus of g

2 – Royal Road for variable length representation

The contribution of each block is fixed and so, the fitness $f_{Nb}(g)$ with g having n blocks is simply :

$$f_{Nb}(g) = \frac{1}{N} \sum_{i=1}^{N} B_b(g, l_i) = \frac{n}{N}$$

To efficiently reach an optimum, the EA system has to create and combine blocks without breaking existing structures

2.2 – Example with $\Sigma = \{A, T, G, C\}$ *and* $b = 3$
$g \in G_\Sigma$ having 2 blocks :

$$g = AAGGTAGGGTAATTTCACTGGG \ , \ f_{Nb}(g) = \frac{2}{4} = 0.5$$

Note that only the presence of a block is taken into account, neither its position nor its repetition

3 – NK-Landscapes

Study link between epistatis and ruggedness of search spaces (Kauffman 1993)

3.1 – Definition

$$f_{NK}(x) = \frac{1}{N} \sum_{i=1}^{N} f_i(x_i; x_{i_1}, \ldots, x_{i_K})$$

- N the size of genotypes
- $K \leq N - 1$ number of epistatic link
- x a genotype of $\{0,1\}^N$
- x_k the k^{th} locus of x
- $\{i_1, \ldots, i_K\} \subset \{1, \ldots, i-1, i+1, \ldots, N\}$
- $f_i : \{0,1\}^{K+1} \rightarrow [0,1)$

3.2 – example of NK-Landscapes : $N = 4,\ K = 2$

$$x = 0110$$

$x_1 x_2 x_4$	f_1
000	0.546
001	0.324
010	0.471
011	0.234
...	...

$x_1 x_2 x_3$	f_2
000	0.216
001	0.654
010	0.156
011	0.215
...	...

$x_2 x_3 x_4$	f_3
000	0.343
...	...
101	0.158
110	0.668
111	0.365

$x_1 x_2 x_4$	f_4
000	0.116
001	0.155
010	0.789
011	0.345
...	...

$$\begin{aligned} f_{NK}(x) &= \tfrac{1}{4}(\ f_1(010) + f_2(011) + f_3(110) + f_4(010)\ ) \\ &= \tfrac{1}{4}(\ 0.471 + 0.215 + 0.668 + 0.789\ ) \\ &= 0.53575 \end{aligned}$$

3.2 – example of NK-Landscapes : $N = 4,\ K = 2$

$$x = 0110$$

$x_1x_2x_4$	f_1
000	0.546
001	0.324
010	0.471
011	0.234
...	...

$x_1x_2x_3$	f_2
000	0.216
001	0.654
010	0.156
011	0.215
...	...

$x_2x_3x_4$	f_3
000	0.343
...	...
101	0.158
110	0.668
111	0.365

$x_1x_2x_4$	f_4
000	0.116
001	0.155
010	0.789
011	0.345
...	...

$$\begin{aligned} f_{NK}(x) &= \frac{1}{4}(f_1(010) + f_2(011) + f_3(110) + f_4(010)) \\ &= \frac{1}{4}(0.471 + 0.215 + 0.668 + 0.789) \\ &= 0.53575 \end{aligned}$$

3.2 – example of NK-Landscapes : $N = 4,\ K = 2$

$$x = 0110$$

$x_1x_2x_4$	f_1
000	0.546
001	0.324
010	0.471
011	0.234
...	...

$x_1x_2x_3$	f_2
000	0.216
001	0.654
010	0.156
011	0.215
...	...

$x_2x_3x_4$	f_3
000	0.343
...	...
101	0.158
110	0.668
111	0.365

$x_1x_2x_4$	f_4
000	0.116
001	0.155
010	0.789
011	0.345
...	...

$$\begin{aligned} f_{NK}(x) &= \tfrac{1}{4}\left(f_1(010) + f_2(011) + f_3(110) + f_4(010) \right) \\ &= \tfrac{1}{4}\left(0.471 + 0.215 + 0.668 + 0.789 \right) \\ &= 0.53575 \end{aligned}$$

3.2 – example of NK-Landscapes : $N = 4,\ K = 2$

$$x = 0110$$

$x_1x_2x_4$	f_1
000	0.546
001	0.324
010	0.471
011	0.234
...	...

$x_1x_2x_3$	f_2
000	0.216
001	0.654
010	0.156
011	0.215
...	...

$x_2x_3x_4$	f_3
000	0.343
...	...
101	0.158
110	0.668
111	0.365

$x_1x_2x_4$	f_4
000	0.116
001	0.155
010	0.789
011	0.345
...	...

$$\begin{aligned} f_{NK}(x) &= \tfrac{1}{4}(\ f_1(010) + f_2(011) + f_3(110) + f_4(010)\ ) \\ &= \tfrac{1}{4}(\ 0.471 + 0.215 + 0.668 + 0.789\ ) \\ &= 0.53575 \end{aligned}$$

3.2 – example of NK-Landscapes : $N = 4,\ K = 2$

$$x = 0110$$

$x_1x_2x_4$	f_1
000	0.546
001	0.324
010	0.471
011	0.234
...	...

$x_1x_2x_3$	f_2
000	0.216
001	0.654
010	0.156
011	0.215
...	...

$x_2x_3x_4$	f_3
000	0.343
...	...
101	0.158
110	0.668
111	0.365

$x_1x_2x_4$	f_4
000	0.116
001	0.155
010	0.789
011	0.345
...	...

$$\begin{aligned} f_{NK}(x) &= \tfrac{1}{4}\left(f_1(010) + f_2(011) + f_3(110) + f_4(010) \right) \\ &= \tfrac{1}{4}\left(0.471 + 0.215 + 0.668 + 0.789 \right) \\ &= 0.53575 \end{aligned}$$

4 – Epistatic Road for variable length representation

- Genes are defined as blocks : exactly as in RR
- The contribution of a gene depends on the presence of others : exactly as in NK-landscapes

4.1 – Definition

g is a genotype of G_Σ

$$f_{NKb}(g) = \frac{1}{N} \sum_{i=1}^{N} f_i(B_b(g, l_i); B_b(g, l_{i_1}), \ldots, B_b(g, l_{i_K}))$$

An ER-landscape depends on the three parameters N, K and b

- N : size of alphabet and number of contribution
- b : size of block
- K : number of epistatic link

4 – Epistatic Road for variable length representation

4.2 – Example with $N = 4$, $K = 2$, $b = 3$, $\Sigma = \{A, T, G, C\}$

$$g = AAGGTAGGGTAATTTCACTGGG$$

so we have :

$$x = 0110$$

and then we have :

$$f_{NKb}(g) = 0.53575$$

Note :

- For $K = 0$: closed to the corresponding Royal Road for VLR
- For $K = N - 1$: the majority of the roads leads to local optima

5 – Fitness landscape analysis

5.1 – Random walks, autocorrelation function and correlation length

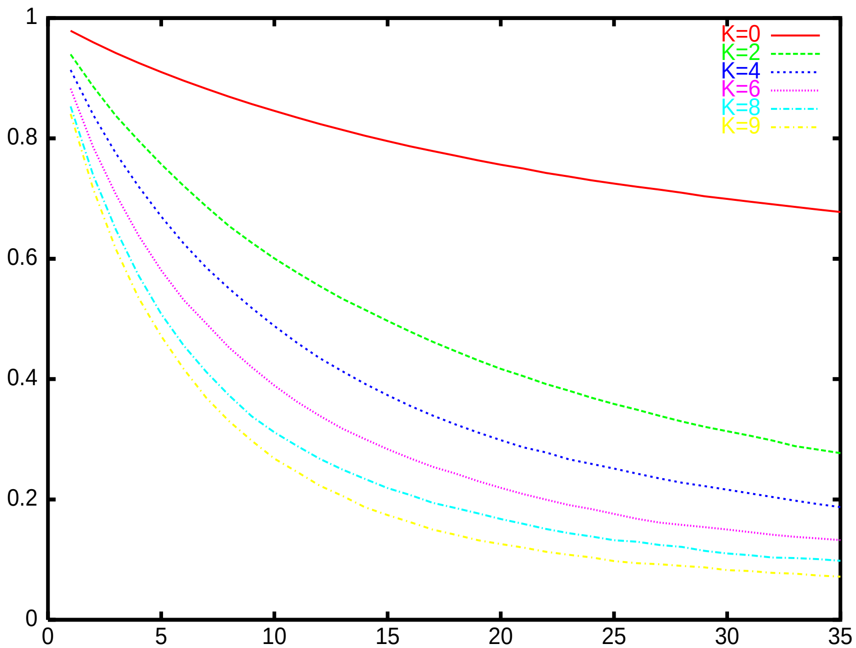

Figure 1: Autocorrelation function of ER for $N = 10$ $b = 2$

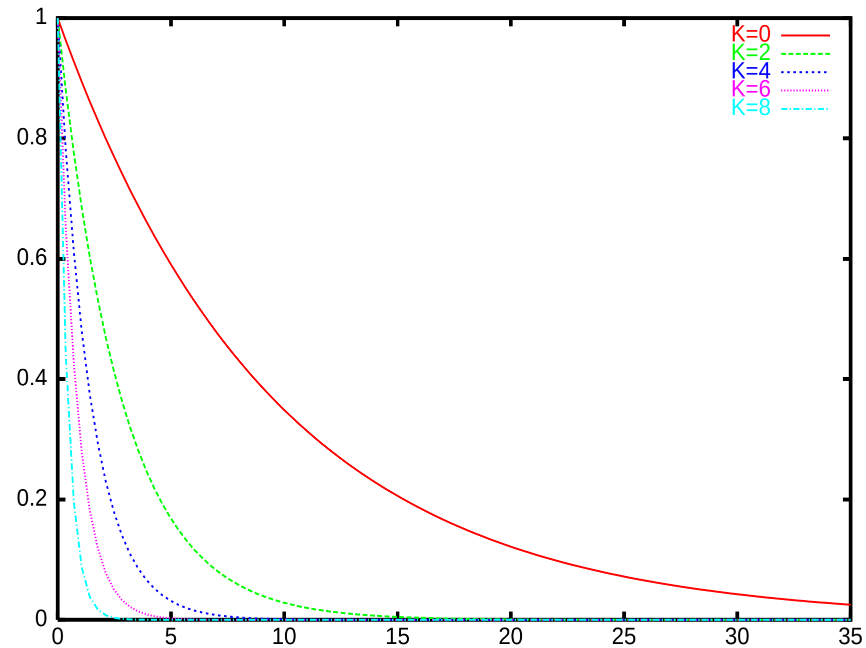

Figure 2: Theoritical autocorrelation function of NK for $N = 10$

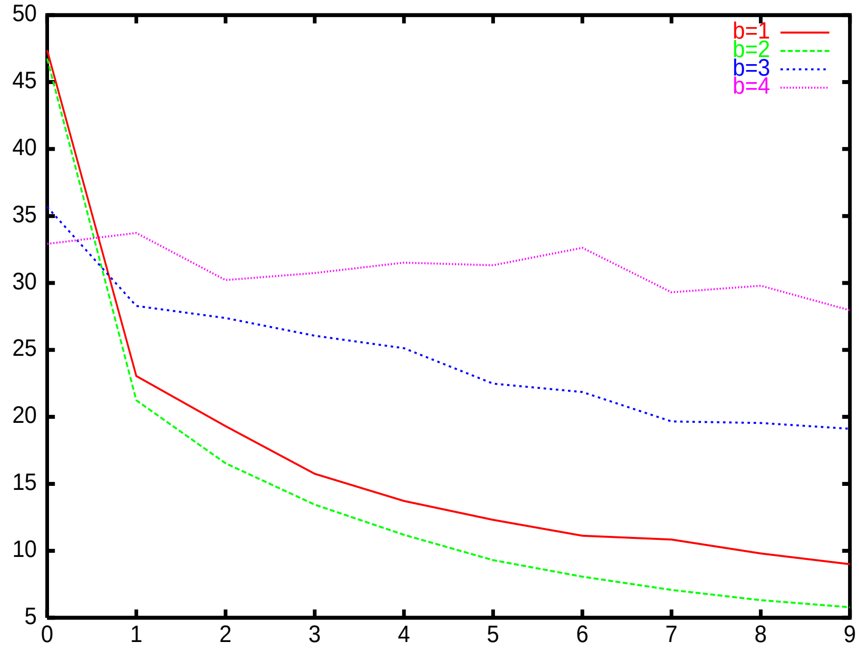

Figure 3: Mean correlation length as a function of K on ER $N = 10$

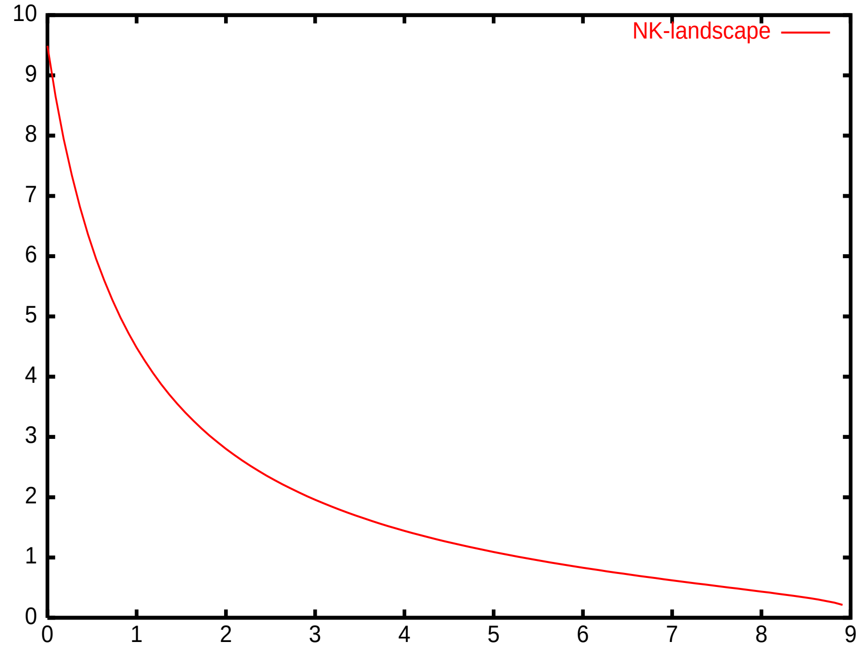

Figure 4: Theoritical correlation length as a function of K on NK $N = 10$

5.2 – Adaptive walks and local optima

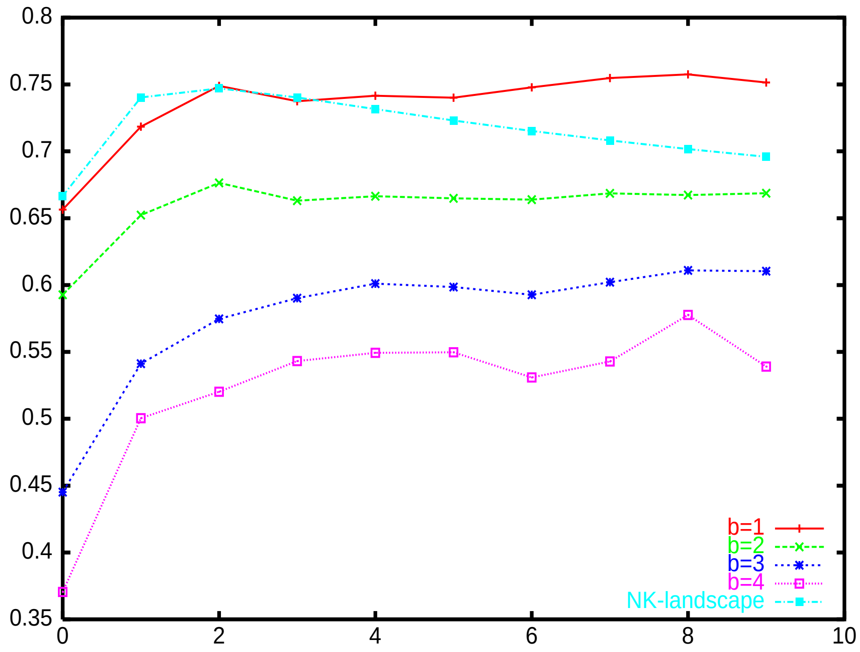

Figure 5: Mean fitness of local optima of ER for $N = 10$

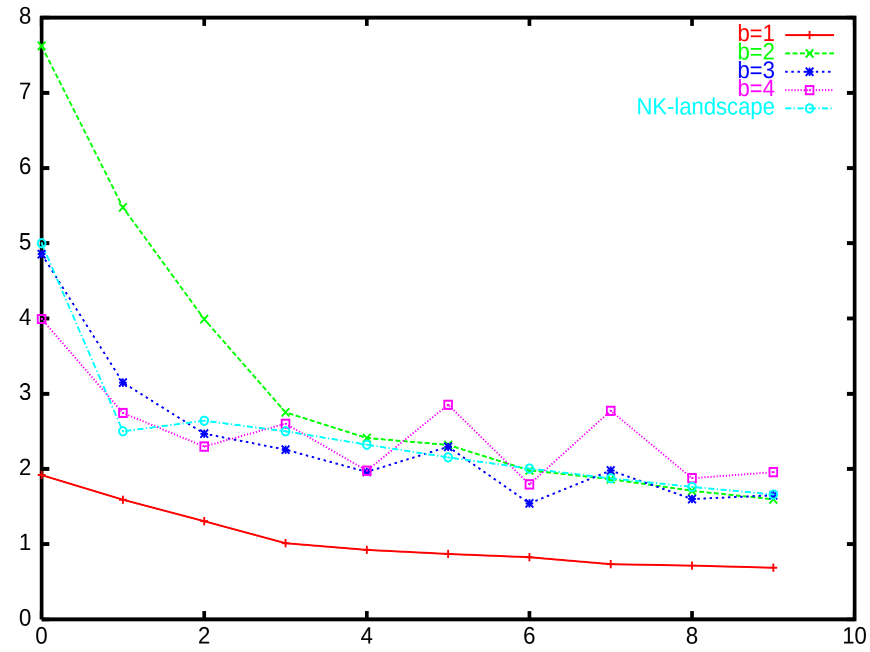

Figure 6: Mean length of adaptive walks on ER for $N = 10$

5.3 – Neutrality

Block size	$N = 8,\ K = 4$		
	Lower	Equal	Higher
$b = 2$	7.2	85.8	7.0
$b = 3$	2.8	94.4	2.8
$b = 4$	0.5	98.9	0.6

Proportion of Lower, Equal and Higher neighbour

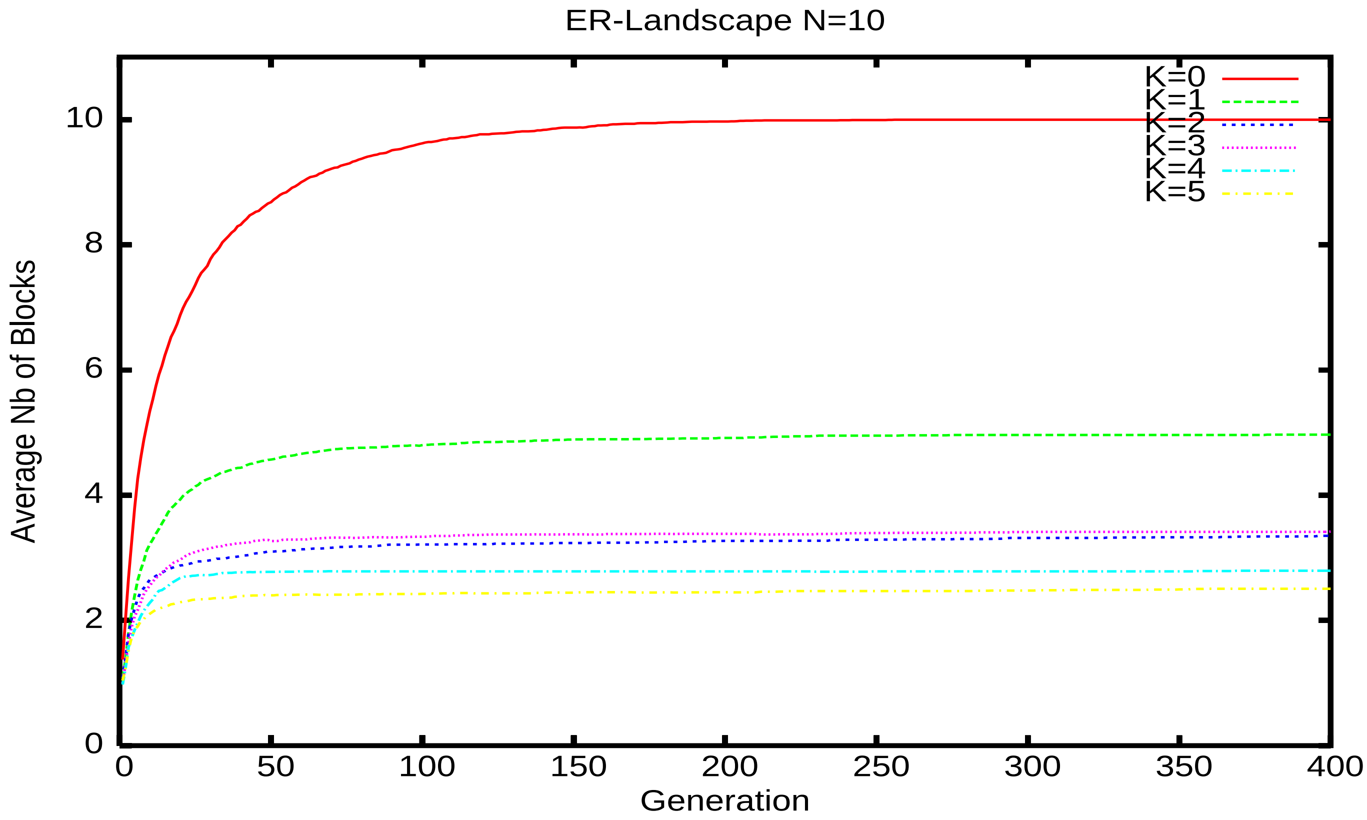

Figure 7: Evolution of average number of blocks on ER N=10 b=4

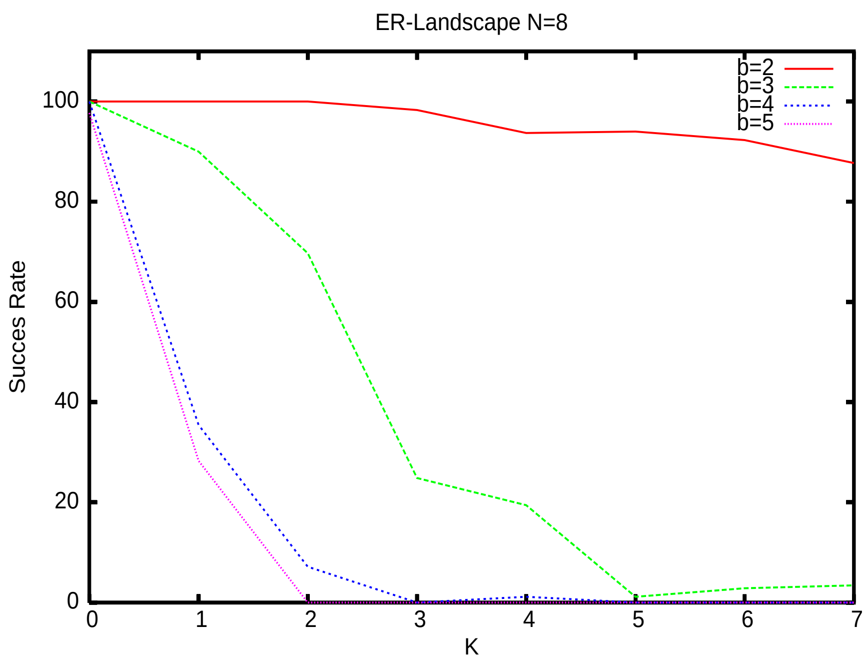

Figure 8: Success rate as a function of K on ER N=8

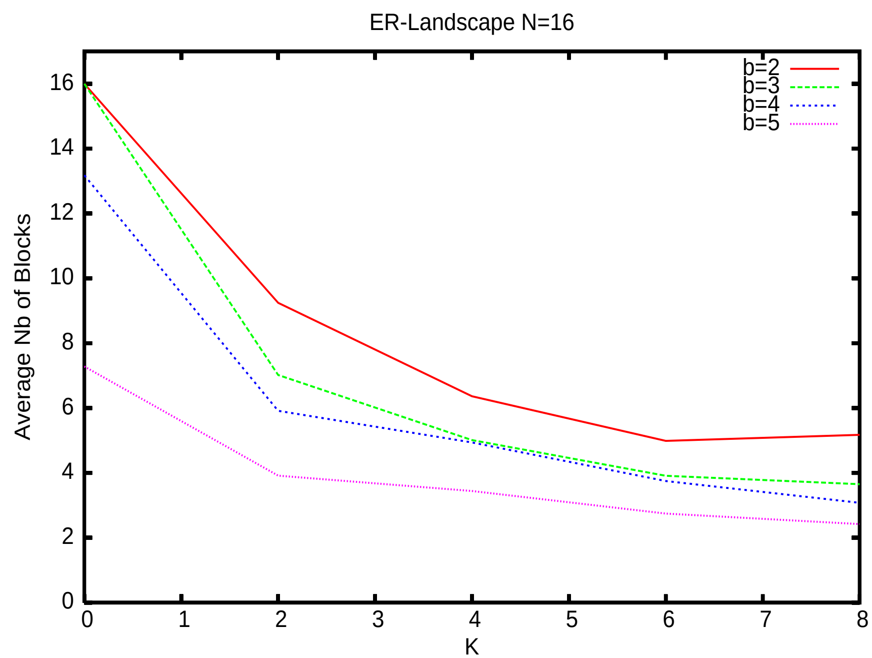

Figure 9: Average nb of blocks as a function of K on ER N=16

7 – Conclusion

- Epistatic Road landscapes : neutrality and ruggedness can be tuned
- Statistical measures : b controls the neutrality and K controls the ruggedness
- Experiments : expected difficulty according to parameters b and K

$\Rightarrow$ A ready-to-use VLR problem of tunable difficulty is available

$\Rightarrow$ Difficulty in variable length representations for Evolutionary Algorithms