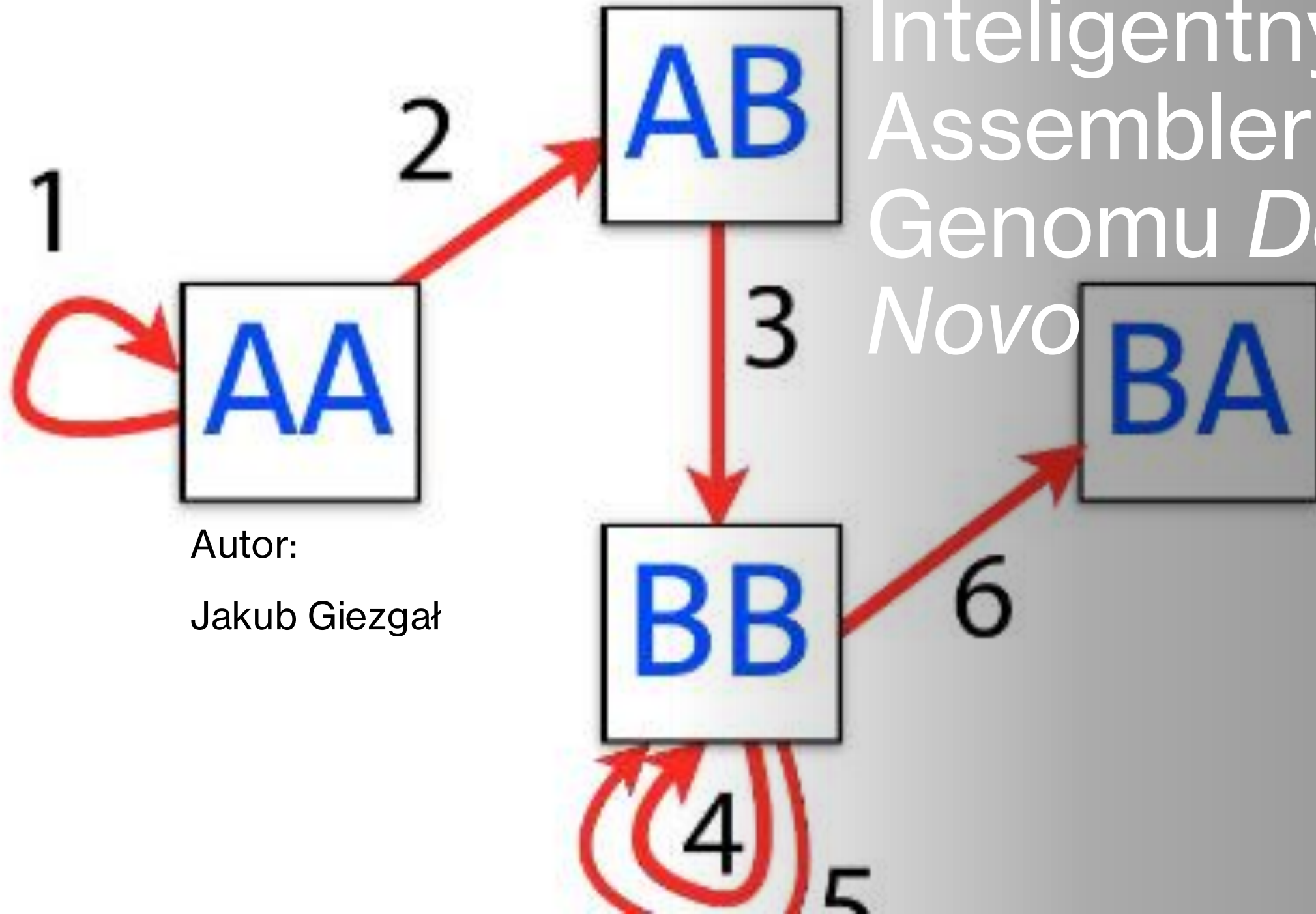
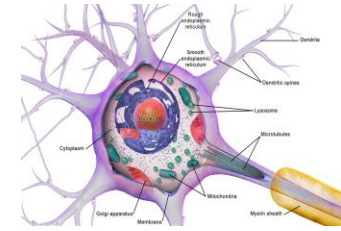


Inteligentny Assembler Genomu *De Novo*



Inteligently \Leftrightarrow Auto-Tuning



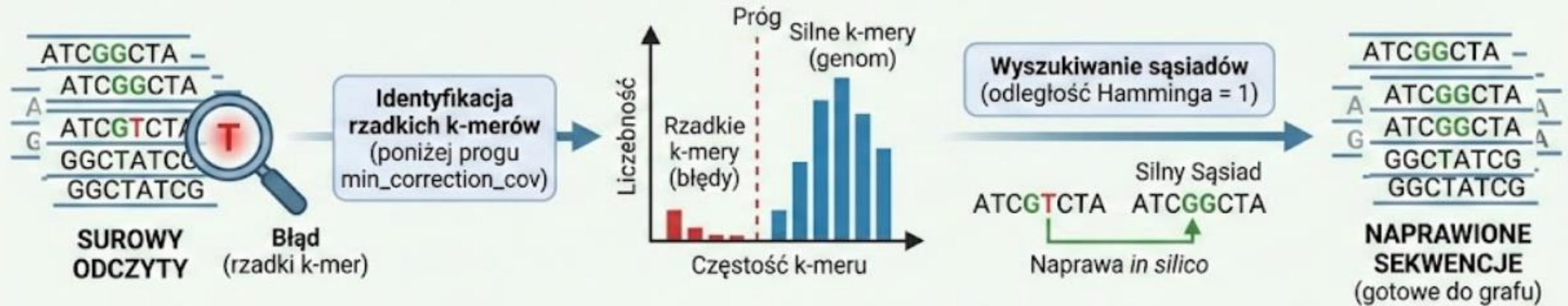
Condition	Mode	Parameters	Goal
$\text{EstCov} \geq 7.5$	STABLE HIGH	$K=21$, $\text{MinCov}=2$	High precision for dense data. Noise filtering.
$4.0 \leq \text{EstCov} < 7.5$ ($\text{ErrorRatio} < 25\%$)	PRECISION RESCUE	$K=21$, $\text{MinCov}=1$	Preserves sparse reads, large K for precision.
$4.0 \leq \text{Cov} < 7.5$ ($\text{ErrorRatio} \geq 25\%$)	AGGRESSIVE FILTER	$K=17$, $\text{MinCov}=2$	Smaller K increases overlap probability, $\text{MinCov}=2$ removes noise.
$\text{Cov} < 4.0$	DESPERATE	$K=15$, $\text{MinCov}=1$	Extremely low coverage. Desperate attempt to assemble anything.

Przed rozpoczęciem asemblacji *de novo*, program wykonuje szybki scan odczytów używając małych *k*-merów o długości 15. Dzięki temu wyliczane są dwie metryki od których zależy zastosowana strategia:

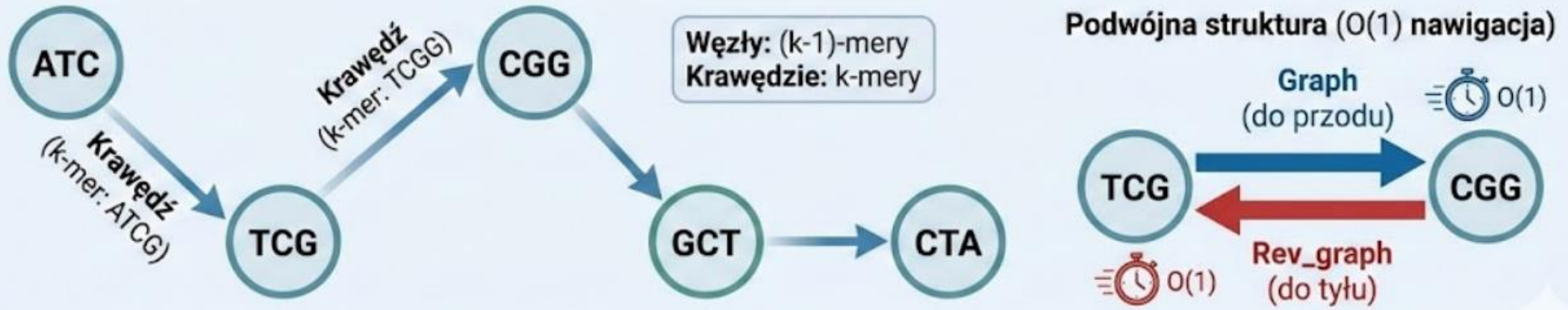
- **EstCov:** Mediana wystąpień *k*-merów występujących więcej niż jeden raz.
- **ErrorRatio:** Stosunek unikalnych *k*-merów do unikalnych sekwencji.

```
+1 for extending by an edge.  
  
+ Edge Weight (prefers paths with high coverage).  
  
-5 penalty for entering a Dead-End.  
  
-0.5 penalty for ambiguity (branching).
```

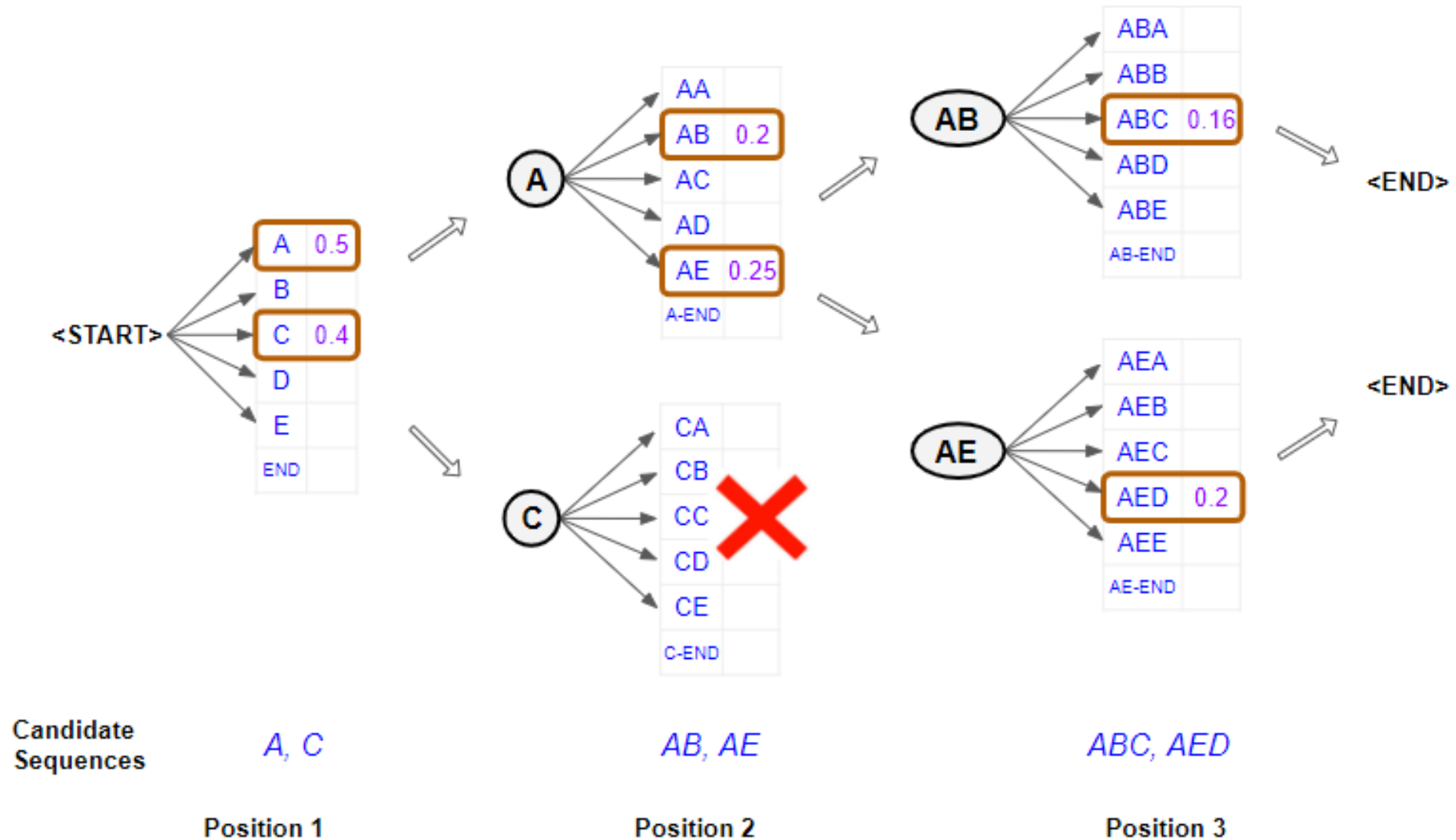
SPECTRAL ERROR CORRECTION



STRUKTURA GRAFU DE BRUIJNA



Beam Search > Greedy





Grid Search Is All You Need!

REFERENCES



PEVZNER, P. A., TANG, H., & WATERMAN, M. S. (2001). "AN EULERIAN PATH APPROACH TO DNA FRAGMENT ASSEMBLY". PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES.



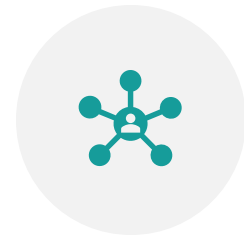
ZERBINO, D. R., & BIRNEY, E. (2008). "VELVET: ALGORITHMS FOR DE NOVO SHORT READ ASSEMBLY USING DE BRUIJN GRAPHS". GENOME RESEARCH.



LOWERRE, B. T. (1976). "THE HARPY SPEECH RECOGNITION SYSTEM. (PH.D. THESIS)". CARNEGIE MELLON UNIVERSITY.



CHIKHI, R., & MEDVEDEV, P. (2014). "INFORMED AND AUTOMATED K-MER SIZE SELECTION FOR GENOME ASSEMBLY". BIOINFORMATICS.



KELLEY, D. R., SCHATZ, M. C., & SALZBERG, S. L. (2010). "QUAKE: QUALITY-AWARE DETECTION AND CORRECTION OF SEQUENCING ERRORS". GENOME BIOLOGY.