Combinatorics of Circular Codes

Serdiukov, Aleksandr (School: School 564)

My work is devoted to studying the combinatorics of 216 maximal C3 circular codes - a special sets of words. Circularity means, that every new word, formed from the elements of this set and written on a circle has at most one decomposition over this set. Circular codes found in genes are suspected to be one of the mechanisms of protection against the reading frame shift in biology. Also, they can be responsible for verifying that the gene code section does not belong to another organism. In 2008 all 216 trinucleotide circular codes were generated. One of the latest studies of circular codes that was conducted in 2014 showed the basic structure of each trinucleotide circular code. I decided to explore the structure of all these codes together in a whole new way using a graph theory. So, I built a special graph which I called “the variant codes graph”. Its nodes are circular codes, and edge between two nodes show that intersection of these two sets of words is maximal (18 words). I found that the maximal group of letter mappings, which preserve the circularity property, is D4 over the genetic alphabet. Also, I found that degrees of D4-orbits in the factor-graph coincide with degrees of their members in source graph – that does not hold for an arbitrary graph and it is a very special property which wasn't known before. Moreover, I found an invariant - multiset of the neighbors’ degrees, which differs 20 out of 27 D4-orbits. This result is another significant step in describing the structure of all trinucleotide circular codes and finding a way to work with them. It showed that it is very effective to work with them using graph theory

Awards Won:

Third Award of $1,000