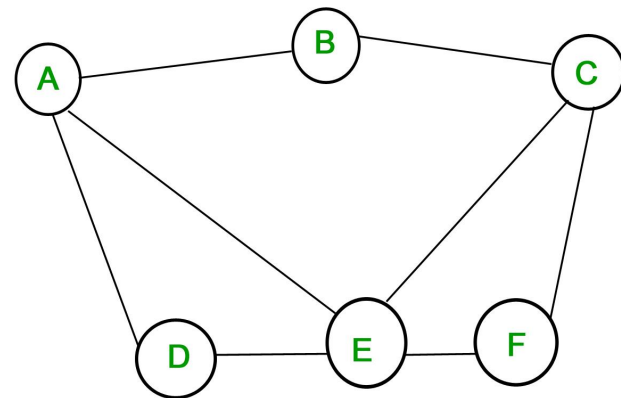
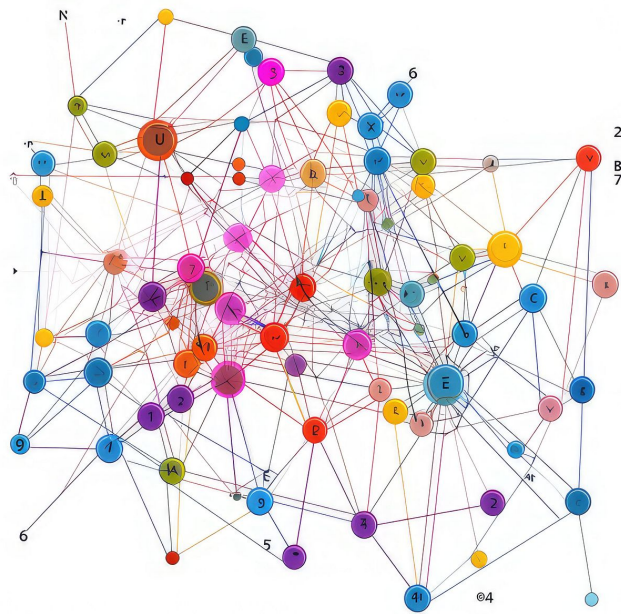


Montaje de secuencias

Grafos o gráficas

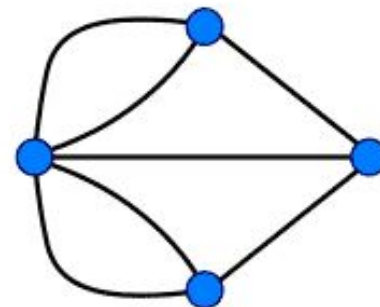
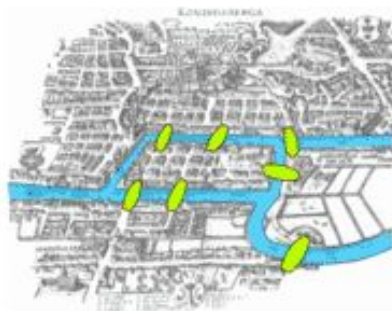


¿Qué es un grafo?

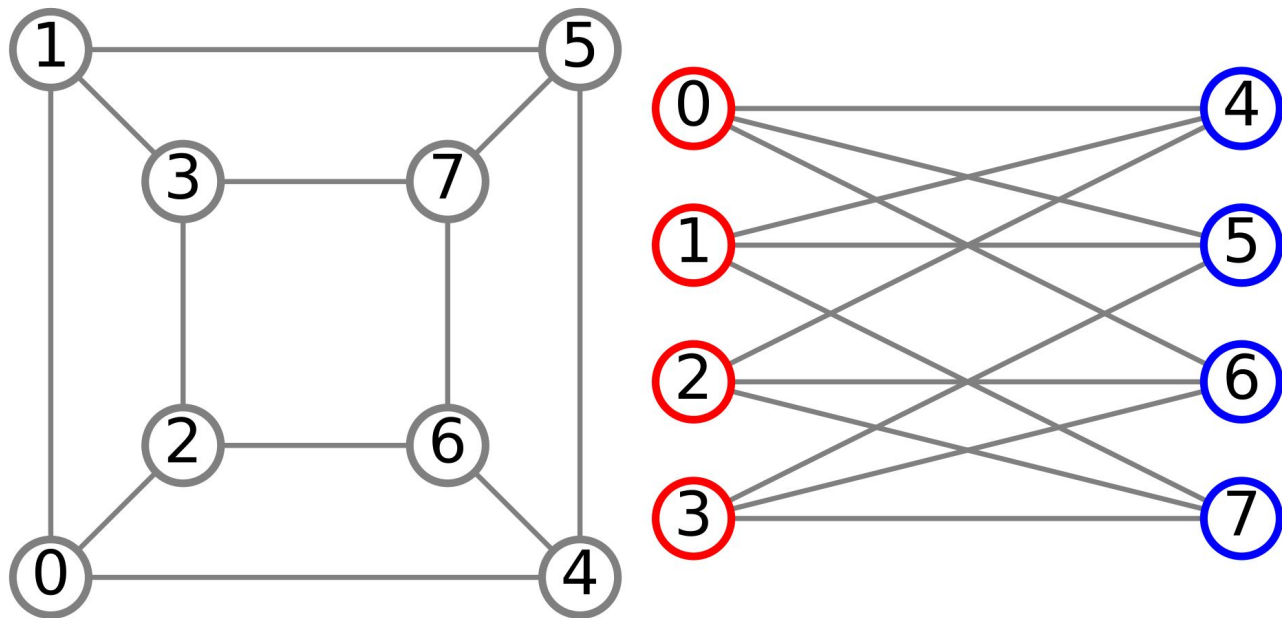
Un **grafo no dirigido** $G(V,E)$ es una pareja ordenada en donde V es un conjunto no vacío de vértices y E es el conjunto de aristas, el cual consta de pares **no ordenados** de vértices, como $\{x,y\}$.

Un **grafo dirigido** $G(V,E)$ es una pareja ordenada en donde V es un conjunto no vacío de vértices y E es el conjunto de aristas, el cual consta de pares **ordenados** de vértices, como (x,y) .

Puentes de Königsberg

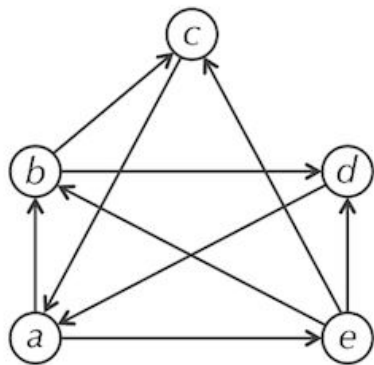


Isomorfismo de grafos



Representaciones de grafos

Graph



Adjacency Matrix

	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>
<i>a</i>	0	1	0	0	1
<i>b</i>	0	0	1	1	0
<i>c</i>	1	0	0	0	0
<i>d</i>	1	0	0	0	0
<i>e</i>	0	1	1	1	0

Adjacency List

a is adjacent to *b* and *e*

b is adjacent to *c* and *d*

c is adjacent to *a*

d is adjacent to *a*

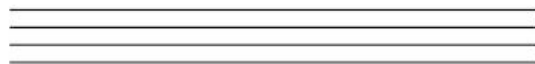
e is adjacent to *b*, *c*, and *d*

Secuenciación de ADN

Un investigador toma una pequeña muestra de tejido que contiene millones de células con ADN idéntico, se utilizan métodos bioquímicos para quebrar el ADN en fragmentos y luego se secuencian estos fragmentos para generar los “reads”.

La tarea de utilizar estos fragmentos para reconstruir la cadena original se le conoce como montaje de secuencias.

Multiple identical
copies of a genome



Shatter the genome
into reads



Sequence the reads

AGAATATCA

TGAGAATAT

GAGAATATC

Assemble the
genome using
overlapping reads

AGAATATCA

GAGAATATC

TGAGAATAT

...TGAGAATATCA...

Algunos retos

- No sabemos a priori que hebra estamos analizando en cada read
- Las máquinas modernas de secuenciación no son perfectas
- Algunas regiones del genoma no van a ser cubiertas por algún read

En este caso, vamos a asumir que no existen errores y que los métodos modernos alcanzan a secuenciar todo el genoma.

Composición de k-meros

Dada una cadena texto, la composición de k-meros es la colección de todas las subcadenas k-meros en dicha cadena.

$$\textit{Composition}_3(\text{TATGGGGTGC}) = \{\text{ATG}, \text{GGG}, \text{GGG}, \text{GGT}, \text{GTG}, \text{TAT}, \text{TGC}, \text{TGG}\}.$$

Reconstrucción de una cadena

AAT ATG GTT TAA TGT

TAA

AAT

ATG

TGT

GTT

TAATGTT

Una reconstrucción más larga

AAT ATG ATG ATG CAT CCA GAT GCC GGA GGG GTT TAA TGC TGG TGT

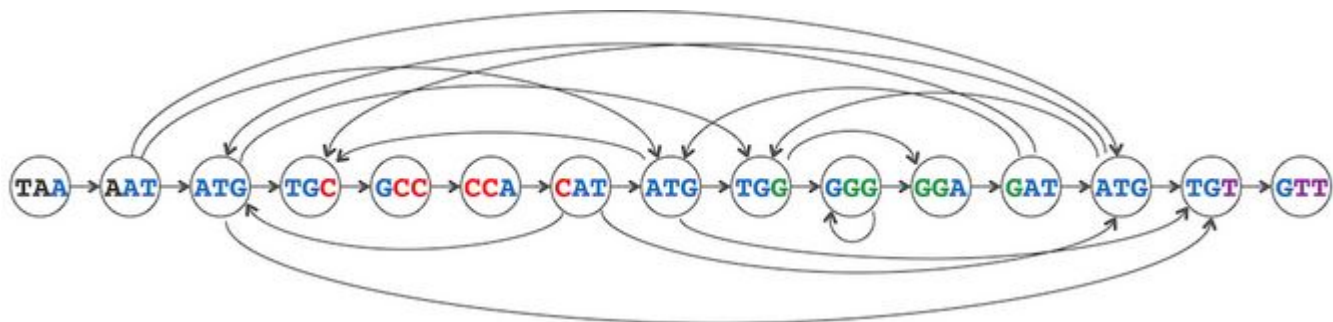
TAA
AAT
ATG
TGC
GCC
CCA
CAT
ATG
TGG
GGA
GAT
ATG
TGT
GTT

Utilizamos grafos

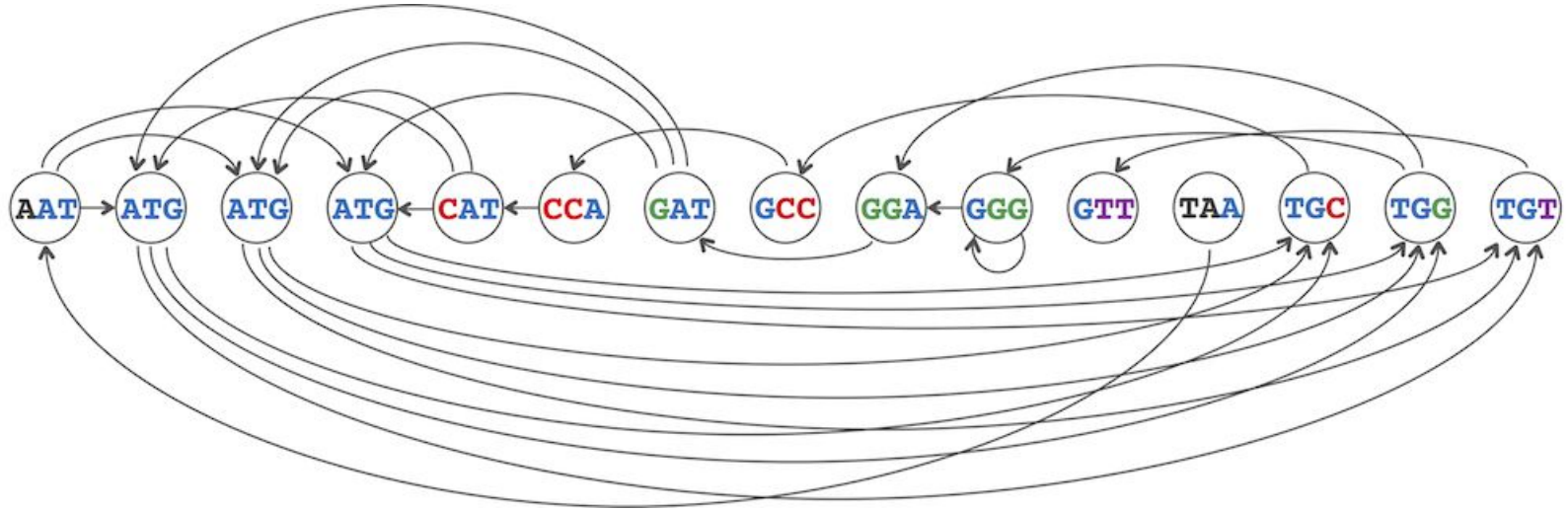
TAA
AAT
ATG
TGC
GCC
CCA
CAT
ATG
TGG
GGG
GGA
GAT
ATG
TGT
GTT
TAATGCCATGGGATGTT



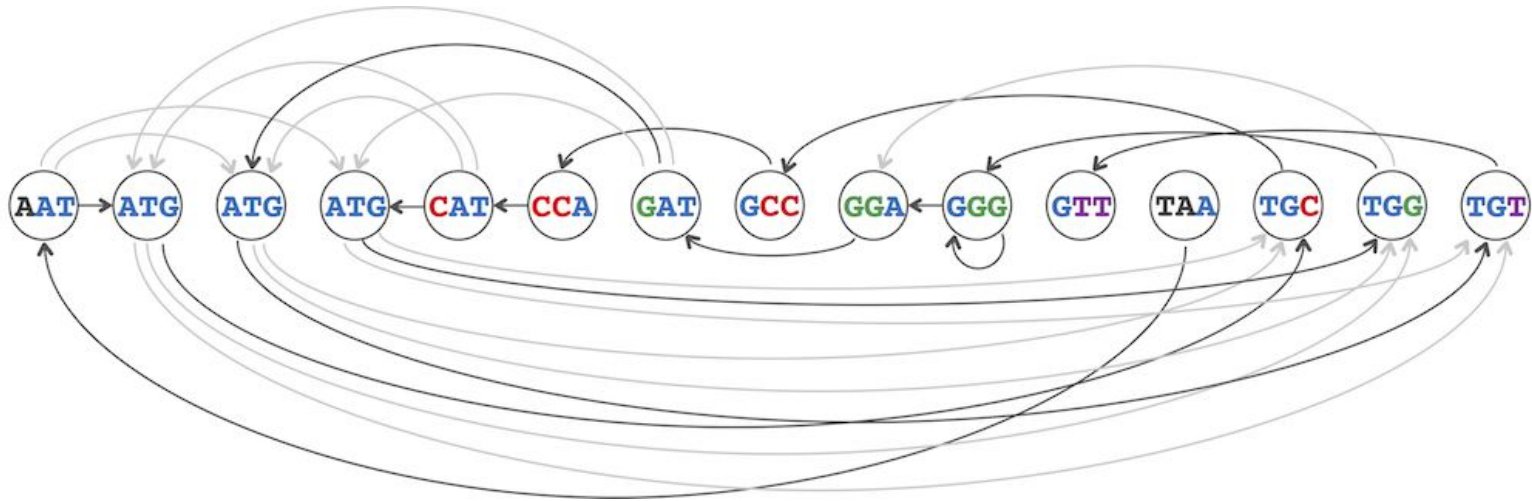
Grafo dirigido



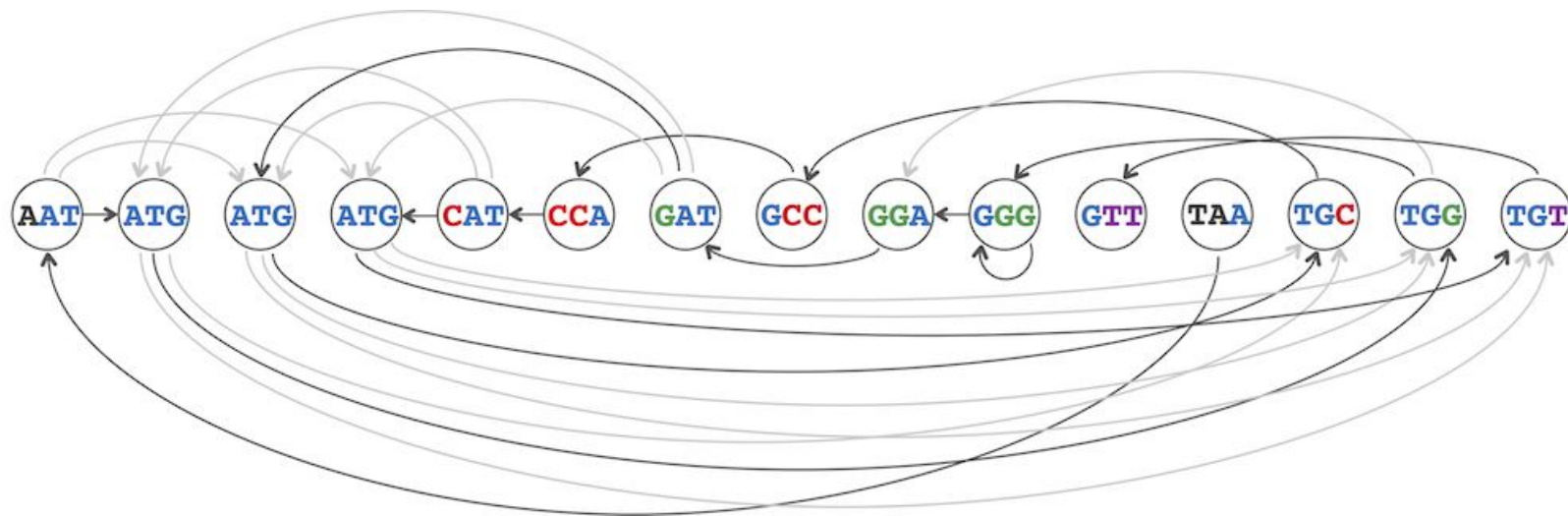
Si desordenamos un poco...



Si desordenamos un poco...

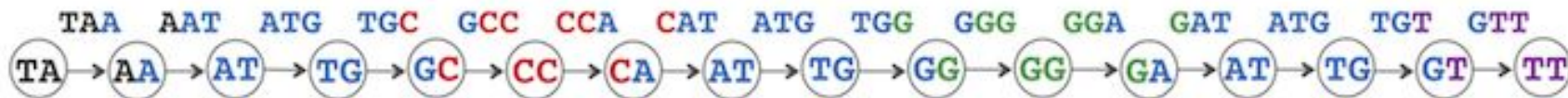
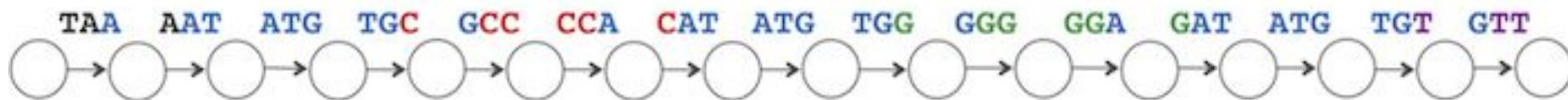


Camino Hamiltoniano

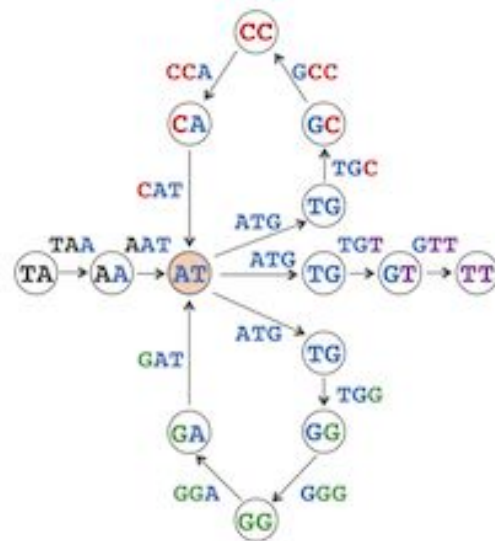
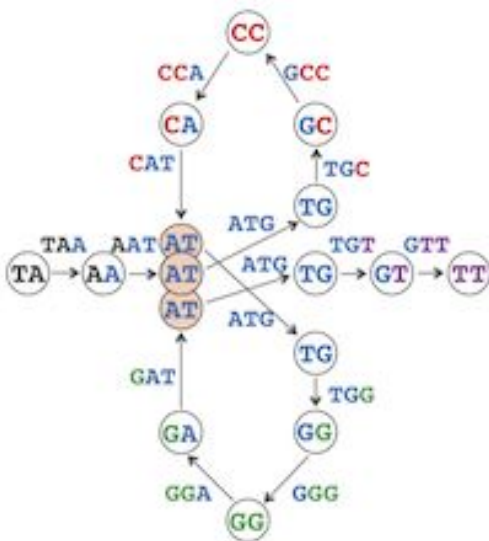
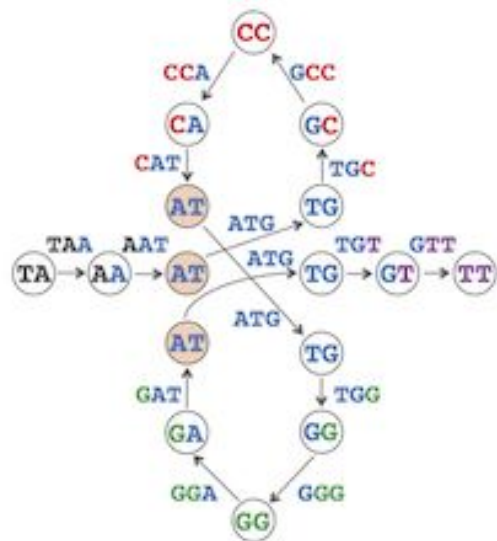


Otra forma de construir grafos

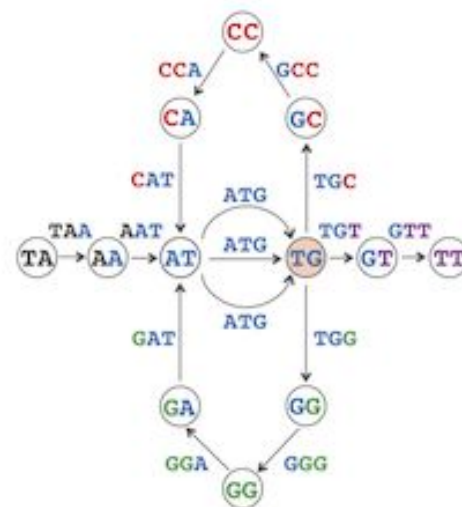
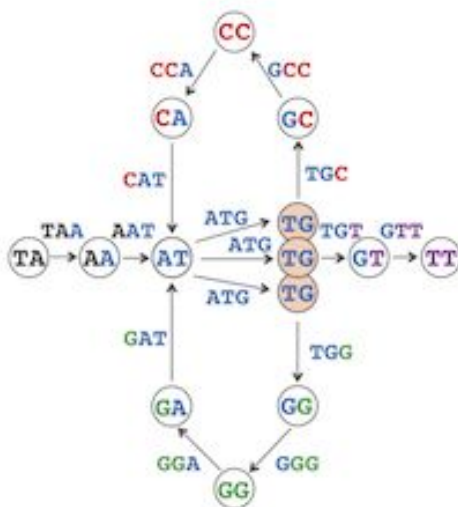
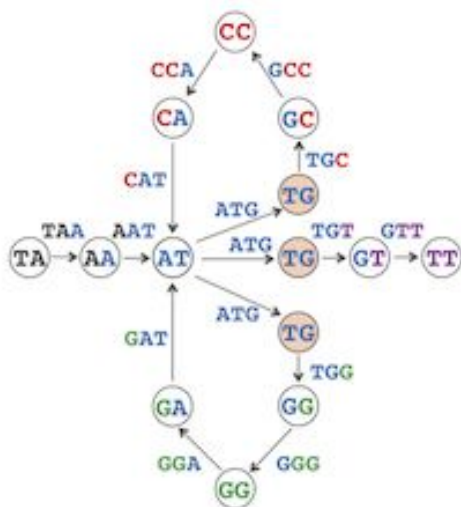
TAA AAT ATG TGC GCC CCA CAT ATG TGG GGG GGA GAT ATG TGT GTT



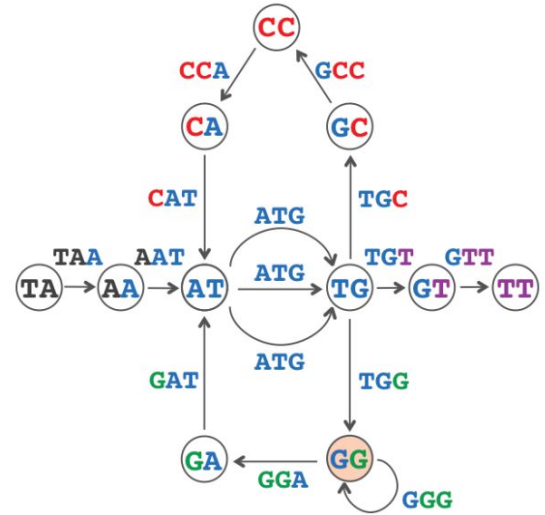
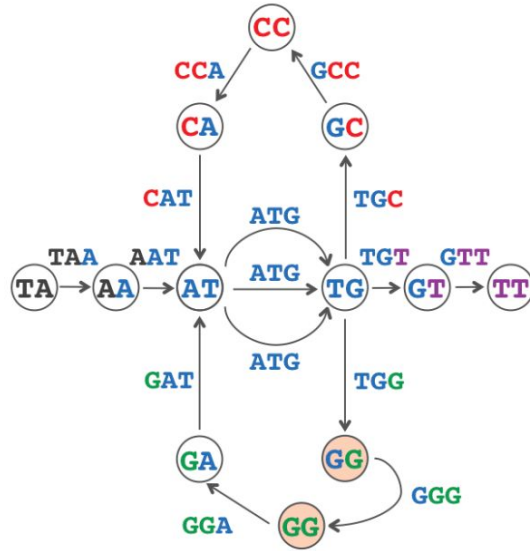
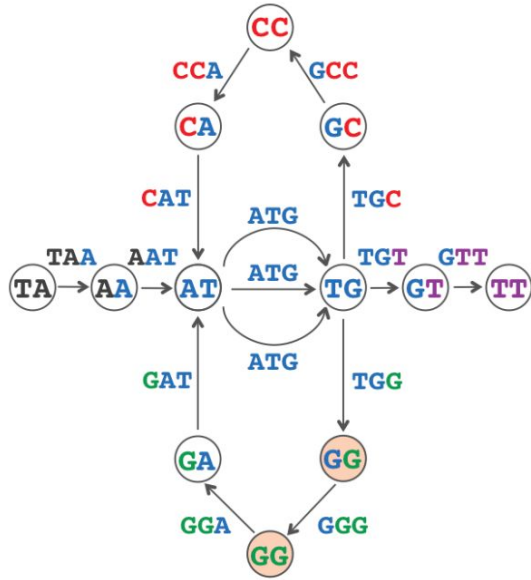
Transformamos el grafo



Transformamos el grafo



Transformamos el grafo



Caminos eulerianos

“Un camino o recorrido por un grafo que usa todas las aristas solo una vez”

