



به نام خدا

تاریخ: ۹۹/۴/۱۸

زمان: ۱۱۵ دقیقه

شروع: ۱۰:۳۵

خاتمه: ۱۲:۳۰

آزمون نهایی درس الگوریتم‌های بیوانفورماتیک

دانشگاه کردستان

نکات مهم (حتماً پیش از شروع به پاسخ‌گویی ملاحظه فرمایید)

- پاسخ‌ها را در برگه سفید و به صورت خوانا بنویسید و به صورت واضح عکس بگیرید و در ساعت ۱۲:۳۰ دقیقه در یک ایمیل به آدرس ایمیل uokhomeworks@gmail.com ارسال نمایید. اسم و شماره دانشجویی را در بالای برگه مرقوم فرمایید. ده دقیقه برای ارسال در نظر گرفته شده است. یعنی به ایمیل‌های ارسالی پس از زمان ۱۲:۴۰ دقیقه ترتیب اثر داده نخواهد شد.
- اگر در پرسشی دارای ابهام بودید از طریق پیام خصوصی در سامانه lms دانشگاه سوال کنید.
- دانشجویان لازم است (فردا) ۱۹ تیر ۹۹ از ساعت ۱۰ صبح به ترتیب جدول زیر، از طریق نرم‌افزار اسکایپ با شناسه `sadegh_sulaimany` تماس بگیرند. نمرات تنها به شرط مصاحبه صوتی (و در صورت نیاز تصویری) نهایی خواهند شد. در صورتی که اسم شما در لیست نیست، یا در صورت بروز هر گونه مشکل در آن روز، با شماره همراه ۰۹۰۲۱۰۶۷۳۹۷ تماس بگیرید.

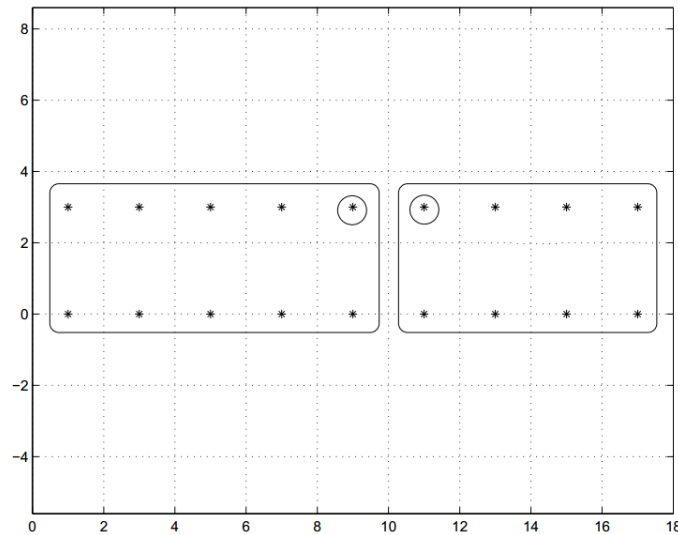
زمان مصاحبه صوتی تصویری در ۱۹ تیر ۹۹		
نوبت	مشخصات دانشجو	ساعت
۱	سودا زارعی	۱۰:۰۰
۲	الهام امیری	۱۰:۱۰
۳	المیرا جمال‌پور	۱۰:۲۰
۴	مسلم یزدان‌پناه	۱۰:۳۰
۵	پدرام فرزین	۱۰:۴۰
۶	شهریار علی‌کرمی	۱۰:۵۰
۷	سوسن سیف‌پناهی	۱۱:۰۰
۸	صادق جعفری	۱۱:۱۰
۹	رحمان مام‌کریمی	۱۱:۲۰
۱۰	رامین سلیمانی	۱۱:۳۰

پرسش ۱: (۳ نمره)

از برنامه‌نویسی پویا برای هم‌تراز کردن دو رشته CATG و CTGCAT استفاده کنید. ماتریس را پر کنید و هم‌ترازی بهینه را بنویسید. امتیاز تطابق را ۲، عدم تطابق را ۱- و گپ را ۱- در نظر بگیرید.

پرسش ۲: (۲ نمره)

شکل زیر را در نظر بگیرید که مشتمل بر دو سطر از نقاط با فواصل مساوی از هم است. اگر از خوشه‌بندی k -means که در اینجا ($k=2$) است برای خوشه‌بندی نقاط استفاده کنیم و دو نقطه ابتدایی را به صورت تصادفی $(9,3)$ و $(11,3)$ انتخاب کنیم، خوشه‌های نهایی را پس از رسیدن به همگرایی (convergence)، تعیین و ترسیم کنید و مرکز نهایی برای هر خوشه را نیز مشخص نمایید. محاسبات را بنویسید.



پرسش ۳: (۲ نمره)

الگوریتم UPGMA را به ماتریس زیر اعمال کنید و درخت حاصل را ترسیم نمایید. محاسبات را بنویسید.

	A	B	C	D
A	0	10	6	4
B	10	0	12	8
C	6	12	0	6
D	4	8	6	0

پرسش ۴: (۲ نمره)

کوتاهترین ابر رشته (Shortest superstring) را برای مجموعه زیر بیابید و گراف متناظر آن را نیز ترسیم کنید.

$$S = \{ATC, CCA, CAG, TCC, AGT\}$$

پرسش ۵: (۲ نمره)

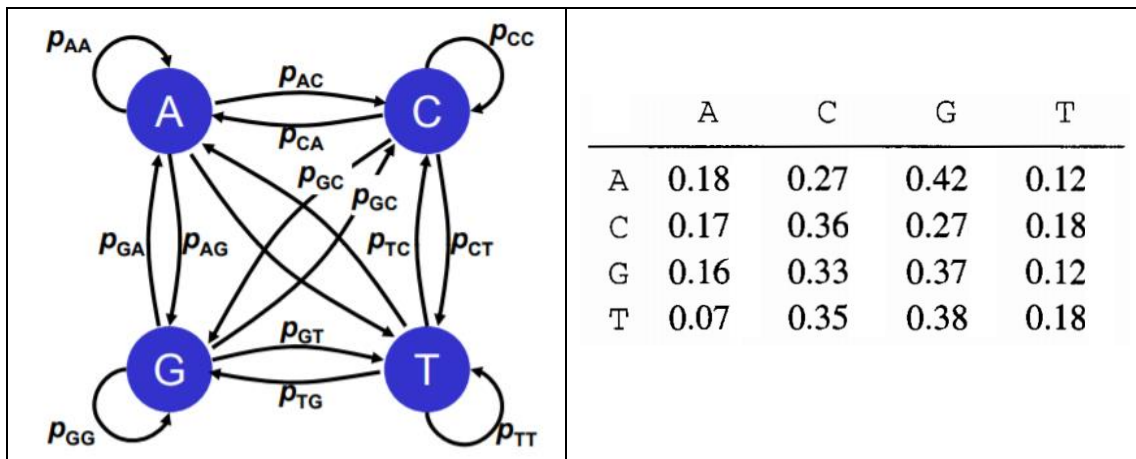
از بیشینه‌سازی زوج‌ها (base pair maximization) برای یافتن ساختار نوع دوم دنباله RNA زیر استفاده کنید و جزییات محاسبه را نشان دهید (می‌توانید برای سهولت، bifurcation را لحاظ نکنید).

CGAGUCG

پرسش ۶: (۱ نمره)

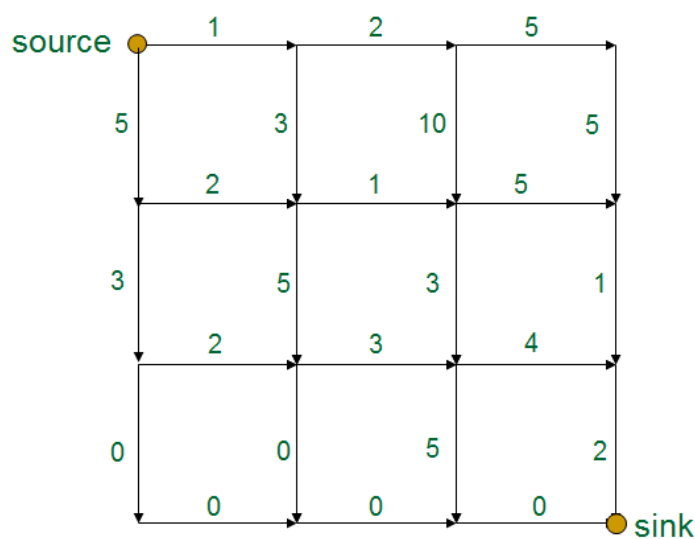
اگر نمودار زیر بیانگر احتمال رویداد هر یک از اجزای DNA پس از دیگری باشد (مثلا P_{AC} به معنای احتمال رویداد C پس از A در رشته باشد و در جدول احتمال نوشته شده به ترتیب ابتدا سطر و سپس ستون باشد، مثلا $P_{AC} = 0.27$ باشد) و زنجیره مارکوف مرتبه ۱ را برای محاسبه احتمال در رشته DNA مدنظر داشته باشیم، کدامیک از دو رشته زیر محتمل تر هستند؟ (احتمال بیشتری بر مبنای زنجیره مارکوف مرتبه یک دارند)

GATCT
CGCGT



پرسش ۷ (نمره مثبت): (۲ نمره)

بر اساس تعریف مسأله توریست منهتن (Manhattan Tourist Problem)، میسر منهتن از Source به Sink را در شبکه زیر بیابید و امتیاز آن را حساب کنید.



با آرزوی سلامتی برای شما