Statistics 251/551 spring 2013
2013: Solutions to sheet 4

This homework will step you through the proof of Lemma 1 in the Jerrum (1995) paper.[1] *I mostly use Jerrum's notation, except that the number of colors is q, not k.*

The setting

The set of available "colors" is $\mathcal{C} = \{1, 2, \ldots, q\}$. We have a graph $\mathcal{G}$ on a finite set of vertices $\mathcal{V} = \{v_1, \ldots, v_n\}$ with edge set $\mathcal{E} = \{e_1, \ldots, e_R\}$. The degree of a vertex v, denoted by $\deg(v)$, is the number of edges having v as one of the endpoints. The neighborhood of vertex v is the set of vertices connected to v by an edge,

$$\mathcal{N}(v) := \{w \in \mathcal{V} : w \neq v \text{ and } \{v, w\} \in \mathcal{E}\}.$$

*A **coloring of the graph** is a map $\sigma : \mathcal{V} \to \mathcal{C}$. The coloring is **proper** if the two vertices that make up each edge are assigned different colors, that is, if*

$$\sigma(v) \notin \{\sigma(w) : w \in \mathcal{N}(v)\} \qquad \text{for every } v \in \mathcal{V}.$$

Denote the set of all proper colorings of the graph by $\mathcal{S}$.

It is easy to see (by means of a greedy coloring method) that $\mathcal{S}$ is nonempty if $q > \Delta$. For $q > 2\Delta$ Jerrum's algorithm generates observations from the uniform distribution π on $\mathcal{S}$ by means of a Markov chain that converges very rapidly towards π.

Transition probabilities

The transition probabilities $P(\sigma, \tau)$ for the chain are defined implicitly by a random method for producing a new coloring τ from a coloring σ.

First define a function TRY that generates a new coloring τ given a new color c, a vertex v, and a current coloring σ. Define $\tau = TRY(c, v, \sigma)$ by

(i) *if $c \in \{\sigma(w) : w \in \mathcal{N}(v)\}$ then $\tau = \sigma$*

(ii) *if $c \notin \{\sigma(w) : w \in \mathcal{N}(v)\}$ then $\tau(v) = c$ and $\tau(w) = \sigma(w)$ for all w not equal to v.*

In other words, $TRY(c, v, \sigma)$ changes the color of v to c, provided the resulting coloring is proper. If the proposed change would create an improper coloring, σ is left unchanged.

---

[1] Available at ***http://onlinelibrary.wiley.com/doi/10.1002/rsa.3240070205/abstract***. You might need to access the site via ***http://sfx.library.yale.edu/sfx_local/azlist*** .

Here is the random procedure corresponding to $P(\sigma, \tau)$:

(a) *Choose a vertex V at random from (the uniform distribution on) $\mathcal{V}$.*

(b) *Choose a color C at random from (the uniform distribution on) $\mathcal{C}$.*

(c) *Define $\tau = TRY(C, V, \sigma)$.*

The coupling

Jerrum's proof works by creating a coupling of a Markov chain $\{Y_t : t \geq 0\}$ with state space $\mathcal{S}$ and an arbitrary (but fixed) initial distribution μ and another Markov chain $\{X_t : t \geq 0\}$ with state space $\mathcal{S}$ and initial distribution π, using a method a little like the coupling used to prove the BLT.

Write info_t *for the information corresponding to everything that has happened up to the completion of step t. Initially $X_0 \sim \pi$ and $Y_0 \sim \mu$, independently. After the completion of step t the X_{t+1} and Y_{t+1} values are coupled as follows*

(a) *Choose a new vertex V_{t+1} at random from (the uniform distribution on) $\mathcal{V}$, independently of* info_t.

(b) *Independently of V_{t+1} and of* info_t, *choose a color C at random from (the uniform distribution on) $\mathcal{C}$.*

(c) *Define $X_{t+1} = TRY(C_{t+1}, V_{t+1}, X_t)$. Based on* info_t *and V_{t+1}, construct a one-to-one function $g : \mathcal{C} \to \mathcal{C}$ then define $Y_{t+1} = TRY(g(C_{t+1}), V_{t+1}, Y_t)$.*

> **Remark.** *The random color $g(C_{t+1})$ is also uniformly distributed on $\mathcal{C}$, independently of* info_t *and V_{t+1}. The change from Y_t to Y_{t+1} still follows the P transition probabilities; marginally, Y is still just a Markov chain with initial distribution μ and transition matrix P.*

If $X_t(V_{t+1}) \neq Y_t(V_{t+1})$ then g is taken to be the identity map (that is, $g(c) = c$ for all c in $\mathcal{C}$) and $Y_{t+1} = TRY(C_{t+1}, V_{t+1}, Y_t)$. The cleverness in Jerrum's algorithm comes from the choice of g for the cases where $X_t(V_{t+1}) = Y_t(V_{t+1})$.

We need some notation for the case where X_t and Y_t agree on the color given to vertex V_{t+1}. After step t of the algorithm the vertex set $\mathcal{V}$ finds itself partitioned into two subsets: $A_t = \{v \in \mathcal{V} : X_t(v) = Y_t(v)\}$, the set of vertices where the two colorings agree; and $D_t = \{v \in \mathcal{V} : X_t(v) \neq Y_t(v)\}$, the set where they disagree.

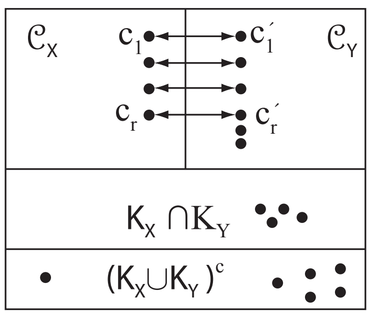
For the sake of notational clarity let me omit some subscripts t and $t+1$, writing V for V_{t+1} and so on. Write $K_X(V)$ for $\{X_t(w) : w \in \mathcal{N}(V)\}$, the set of colors used by X_t for the neighbors of V. Define $K_Y(V)$ analogously. Then $\mathcal{C}_X(V) := K_X(V) \cap K_Y^c(V)$ is the set of colors used only by X_t for the *neighbors of V and $\mathcal{C}_Y(V) := K_Y(V) \cap K_X^c(V)$ is the set of colors used only by Y_t for the neighbors of V.*

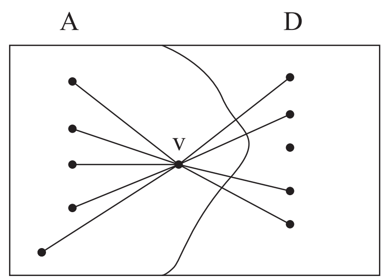
The g is chosen in a way that discourages changes that would turn an agreement at vertex V into a disagreement.

If either $\mathcal{C}_X(V)$ or $\mathcal{C}_Y(V)$ is empty let g be the identity map on $\mathcal{C}$.

If both $\mathcal{C}_X(V)$ and $\mathcal{C}_Y(V)$ are nonempty the smaller of the two sets of colors drives the construction. For concreteness suppose $0 < \#\mathcal{C}_X(V) \leq \#\mathcal{C}_Y(V)$, so that the distinct colors in $\mathcal{C}_X(V)$ can be enumerated as $c_1, c_2, \ldots, c_r$ and the distinct colors in $\mathcal{C}_Y(V)$ can be enumerated as $c'_1, c'_2, \ldots, c'_{r'}$, with $r \leq r'$. Define g by

$$g(c_i) = c'_i \text{ and } g(c'_i) = c_i \qquad \text{for } i = 1, 2, \ldots, r$$
$$g(c) = c \qquad \text{otherwise}$$





The picture on the left represents the partition of $\mathcal{C}$, with the dependence on $v = V_{t+1}$ omitted, for the case $0 < \#\mathcal{C}_X(v) \leq \#\mathcal{C}_Y(v)$. The picture on the right represents $\mathcal{V}$, showing only the the neighbors of v when $X_t(v) = Y_t(v)$, with t subscripts omitted.

If $\#\mathcal{C}_X(V) > \#\mathcal{C}_Y(V) > 0$, reverse the roles of X and Y in the construction, so that g maps $\mathcal{C}_Y(V)$ onto a subset of $\mathcal{C}_X(V)$.

The meeting of the chains

At the random time $T := \min\{t : D_t = \emptyset\}$ the colorings agree for all vertices. At that time both $\mathcal{C}_X(v)$ and $\mathcal{C}_Y(v)$ are empty for every v in $\mathcal{V}$. Subsequently, the algorithm merely modifies the color scheme without creating any new disagreements.

The details of Jerrum's argument appear in the following Problems.

References

Jerrum, M. (1995). A very simple algorithm for estimating the number of k-colorings of a low-degree graph. Random Structures and Algorithms 7*(2), 157–165.*

Homework problems

Write f_t for $\#D_t$, the number of vertices where the colorings disagree after completion of step t. The main idea is to show that $\mathbb{E}f_{t+1} \le (1-\alpha)\mathbb{E}f_t$ where

$$\alpha := \frac{q-2\Delta}{nq} > 0,$$

which will produce a rapidly decreasing bound on the total variation distance between π and the distribution of Y_t.

[1] *(5 points) For each σ and τ in $\mathcal{S}$ show that $P(\sigma,\tau) = P(\tau,\sigma)$. Deduce that the uniform distribution π on $\mathcal{S}$ is the stationary distribution.*

If $P(\sigma,\tau) > 0$ with $\sigma \ne \tau$ then there must exist exactly one vertex v for which $\sigma(v) \ne \tau(v)$. Of course neither $\sigma(v)$ nor $\tau(v)$ can be included in the set of colors used by both σ and τ for $\mathcal{N}(v)$. To get from σ to τ the algorithm must select vertex v (probability n^{-1}) and color $\tau(v)$ (probability q^{-1}); to get from τ to σ the algorithm must select vertex v and color $\sigma(v)$. Both pairs of choices have the same probability, $(nq)^{-1}$.

Time reversibility forces the uniform distribution to be the stationary distribution, as in HW3.2(d).

[2] *(5 points) Explain why*

$$\mathbb{TV}_t := \max_{B\subseteq\mathcal{S}} |\mathbb{P}\{Y_t \in B\} - \pi(B)| \le \mathbb{P}\{T > t\}$$

Hint: The argument is similar to the one used for the proof of the BLT.

For each $B \subseteq \mathcal{S}$,

$$\begin{aligned} |\mathbb{P}\{Y_t \in B\} - \pi(B)| &= |\mathbb{P}\{Y_t \in B\} - \mathbb{P}\{X_t \in B\}| \\ &\le |\mathbb{P}\{Y_t \in B, T \le t\} - \mathbb{P}\{X_t \in B, T \le t\}| \\ &\quad + |\mathbb{P}\{Y_t \in B, T > t\} - \mathbb{P}\{X_t \in B, T > t\}| \end{aligned}$$

From time T onwards the two chains stay the same; they give the same color to each vertex. Thus the first term after the inequality equals zero. For the other term note that

$$-\mathbb{P}\{T > t\} \le \mathbb{P}\{Y_t \in B, T > t\} - \mathbb{P}\{X_t \in B, T > t\} \le \mathbb{P}\{T > t\},$$

which gives $|\mathbb{P}\{Y_t \in B, T > t\} - \mathbb{P}\{X_t \in B, T > t\}| \le \mathbb{P}\{T > t\}$, as in the proof of the BLT. Some of you dervived only the weaker bounder $2\mathbb{P}\{T > t\}$. Some of you only derived an upper bound for $\mathbb{P}\{Y_t \in B\} - \pi(B)$. (Actually, when paired with the analogous bound for B replaced by B^c, the one-sided bound also leads to the desired two-sided bound.)

[3] *(10 points) Explain why* $T = \min\{t : f_t = 0\}$

$$\mathbb{P}\{T > t\} \le \mathbb{P}\{f_t \ne 0\} \le \mathbb{E} f_t.$$

Hint: The random variable f_t *takes only nonnegative integer values.*

By definition, $D_t \ne \emptyset$ (implying $f_t > 0$) for $t < T$. And, because $X_t = Y_t$ for all $t \ge T$, the set D_t stays empty from time T onwards. Consequently, $f_t \ne 0$ if and only if $t < T$. The first inequality is actually an equality: $\mathbb{P}\{T > t\} \le \mathbb{P}\{f_t \ne 0\}$. The second inequality follows by taking expectations on both sides of $\mathbb{I}\{f_t \ge 1\} \le f_t$. Some of you failed to notice that $f_t = 0$ for $t \ge T$.

[4] *(5 points) Explain why* $|f_{t+1} - f_t| \le 1$ *always, with* $f_{t+1} \le f_t$ *if* $V_{t+1} \in D_t$ *and* $f_{t+1} \ge f_t$ *if* $V_{t+1} \in A_t$.

Once again abbreviate V_{t+1} to V. At each step the X- and Y-colors can changed only at vertex V. For an increase in the number of disagreements, it must be that $X_t(V) = Y_t(V)$ but $X_{t+1}(V) \ne Y_{t+1}(V)$. For a decrease in the number of disagreements, it must be that $X_t(V) \ne Y_t(V)$ but $X_{t+1}(V) = Y_{t+1}(V)$. Problems [6] and [7] provide more details about when $f_{t+1} = f_t \pm 1$.

[5] *(5 points) For each vertex* v*, define*

$$d_t(v) := \begin{cases} \#(\mathcal{N}(v) \cap A_t) & \text{if } v \in D_t \\ \#(\mathcal{N}(v) \cap D_t) & \text{if } v \in A_t. \end{cases}$$

Equivalently, $d_t(v)$ *is the number of edges (with* v *as one of the endpoints) that join a point in* A_t *to a point in* D_t*. Explain why the total number of edges that join a point in* A_t *to a point in* D_t *equals*

$$m_t := \sum\nolimits_{v \in A_t} d_t(v) = \sum\nolimits_{v \in D_t} d_t(v).$$

The first expression for m_t comes from working through the list of all the vertices v in A_t while counting how many vertices in D_t are connected to each v on the list. The second expression comes from working through the list of all the vertices v in D_t while counting how many vertices in A_t are connected to each v on the list.

[6] *Consider the case where* $V_{t+1} \in A_t$*. Abbreviate* V_{t+1} *to* V *and* C_{t+1} *to* C.

(i) *(5 points) Explain why* $f_{t+1} = f_t$ *if* $\mathcal{C}_X(V) = \mathcal{C}_Y(V) = \emptyset$.

If both $\mathcal{C}_X(V)$ and $\mathcal{C}_Y(V)$ are empty then $K_X(V) = K_Y(V)$, that is, the X- and Y-colorings agree for all the neighbors of V. The color C_{t+1} is either accepted or rejected by both chains. Both X and Y still give the same color to vertex V.

Without loss of generality suppose $\#\mathcal{C}_X(V) \le \#\mathcal{C}_Y(V)$ and $\mathcal{C}_Y(V) \neq \emptyset$ for the rest of this problem. It is possible that $\mathcal{C}_X(V)$ is empty, which would make (ii) a bit easier.

As I explained in class, the "Without loss of generality" is harmless if we already know that $V_{t+1} = v$, for a specific v in A_t. The analysis for the case where $\#\mathcal{C}_X(v) > \#\mathcal{C}_Y(v)$ is essentially the same as for the case $\#\mathcal{C}_Y(v) \ge \#\mathcal{C}_X(v)$: one merely has to intechange the roles of X and Y. However, the way the question was written did suggest that $\#\mathcal{C}_Y(v) \ge \#\mathcal{C}_X(v)$ for every v in A_t, which need not be true.

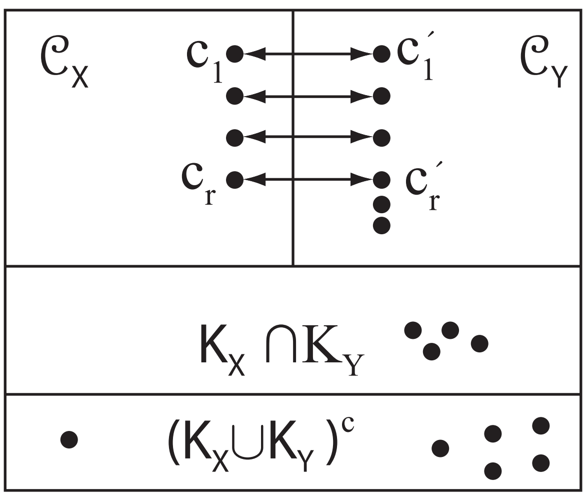
The following three parts should refer to V_{t+1} equal to a (temporarily) fixed v for which $0 < \#\mathcal{C}_X(v) \le \#\mathcal{C}_Y(v)$.

(ii) (15 points) If $C \notin \mathcal{C}_Y(V)$, show that $f_{t+1} = f_t$. Hint: Consider separately the cases where C belongs to $(K_X(V) \cup K_Y(V))^c$ or $K_X(V) \cap K_Y(V)$ or $\mathcal{C}_X(V)$.

If $C \in (K_X(v) \cup K_Y(v))^c$ then neither chain has color C used for any of the neighbors of v and $g(C) = C$, resulting in $X_{t+1}(v) = C = Y_{t+1}(v)$.

If $C \in K_X(v) \cap K_Y(v)$ then $g(C) = C$ and neither chain accepts color C, leaving $X_{t+1}(v) = X_t(v) = Y_t(v) = Y_{t+1}(v)$.

If $C \in \mathcal{C}_X(v)$ then chain X rejects C as one of the colors it already has for a neighbor of v. Chain Y rejects the color $g(C)$, which belongs to $\mathcal{C}_Y(v)$ because of the way the g was constructed.

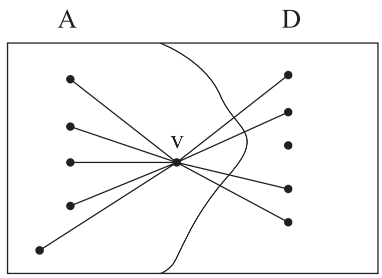


(iii) (5 points) Explain why f_{t+1} might equal $1 + f_t$ if $C \in \mathcal{C}_Y(V)$.

If $C = c_i'$ for some $i \le r$ then $g(C) = c_i \in \mathcal{C}_X(v)$. Both chains accept the new color for vertex v and $f_{t+1} = f_t$. However if $C = c_i'$ for some $r < i \le r'$ then the Y-chain rejects the color C but the X-chain accepts $g(C) = c_i'$ because none of the neighbors of v have X-color c_i'. The result is a different coloring of vertex v by the two chains.

(iv) (5 points) Explain why $\#\mathcal{C}_Y(V) \le d_t(V)$ when $V \in A_t$.

Again argue for V equal to a fixed v in A_t. The colors in $\mathcal{C}_Y(v)$ can appear only on (a subset of the) neighbors of v that lie in D_t. That is, for each color in $\mathcal{C}_Y(v)$ there is at least one edge from v to a neighbor in D_t. The number of colors in $\mathcal{C}_Y(v)$ cannot exceed $d_t(V)$, the total number of edges from v to neighbors in D_t. (A similar argument gives $\#\mathcal{C}_X(v) \le d_t(v)$.)



(v) (10 points) Deduce that

$$\mathbb{P}\{f_{t+1} = 1 + f_t \mid \text{info}_t\} \le \mathbb{P}\{V \in A_t, C \in \mathcal{C}_Y(V) \mid \text{info}_t\} \le \frac{m_t}{nq}.$$

Explain your reasoning in detail.

It is better to argue in the following way. First condition on V_{t+1}. By Problem [4], only if $v \in A_t$ can we have $f_{t+1} = f_t + 1$.

$$\begin{aligned}
&\mathbb{P}\{f_{t+1} = 1 + f_t \mid \text{info}_t\} \\
&\qquad = \mathbb{P}\{f_{t+1} = 1 + f_t, V_{t+1} \in A_t \mid \text{info}_t\} \\
&\qquad = \sum\nolimits_{v \in A_t} \mathbb{P}\{V_{t+1} = v \mid \text{info}_t\} \mathbb{P}\{f_{t+1} = 1 + f_t \mid \text{info}_t, V_{t+1} = v\}
\end{aligned} \tag*{$<1>$}$$

For every v in $\mathcal{V}$ the probability $\mathbb{P}\{V_{t+1} = v \mid \text{info}_t\}$ is equal to n^{-1}. For the analysis of $\mathbb{P}\{f_{t+1} = 1 + f_t \mid \text{info}_t, V_{t+1} = v\}$ suppose, "without loss of generality", that $0 < \#\mathcal{C}_X(v) \le \#\mathcal{C}_Y(v)$. Then

$$\begin{aligned}
&\mathbb{P}\{f_{t+1} = 1 + f_t \mid \text{info}_t, V_{t+1} = v\} \\
&\quad \le \mathbb{P}\{C_{t+1} \in \mathcal{C}_Y(v) \mid \text{info}_t, V_{t+1} = v\} \qquad \text{using without loss of generality} \ldots \\
&\quad \le d_t(v)/q \qquad \text{by (iv).}
\end{aligned}$$

Substitute into <1> to deduce that

$$\mathbb{P}\{f_{t+1} = 1 + f_t \mid \text{info}_t\} \le \sum\nolimits_{v \in A_t} n^{-1} d_t(v)/q = (nq)^{-1} m_t.$$

Some of you claimed that the events $\{V_{t+1} \in A_t\}$ and $\{C_{t+1} \in \mathcal{C}(V_{t+1})\}$ are independent, which is false. Conditioning on $V_{t+1} = v \in A_t$ does let you focus only on the variability in C_{t+1}. Some of you also claimed that the event $\{V_{T+1} \in A_t, C_{t+1} \in \mathcal{C}(V_{t+1})\}$ doesn't depend on info_t, which is also false because the sets A_t and D_t are defined using info_t.

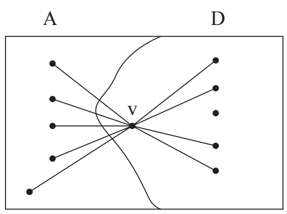
[7] Consider the case where $V_{t+1} \in D_t$. Abbreviate V_{t+1} to V and C_{t+1} to C.

(i) (5 points) For each v in D_t explain why

$$\mathbb{P}(f_{t+1} = f_t - 1 \mid V = v, \text{info}_t) \geq \mathbb{P}(C \notin K_X(v) \cup K_Y(v) \mid V = v, \text{info}_t)$$

If C_{t+1} is in neither the $K_X(v)$ nor the $K_Y(v)$ list then $g(C_{t+1}) = C_{t+1}$ and both chains accept the new color for vertex v. If $v \in D_t$ the number of disagreements is decreased by 1.

(ii) (5 points) Explain why $\#(K_X(v) \cup K_Y(v)) \leq 2\Delta - d_t(v)$ for all v in D_t. Hint: What is the largest number of distinct colors that can be contributed by $\mathcal{N}(v) \cap D_t$?



By definition, $d_t(v) = \#(\mathcal{N}(v) \cap A_t)$ and $\#\mathcal{N}(v) \leq \Delta$. Thus

$$\#(\mathcal{N}(v) \cap D_t) = \#\mathcal{N}(v) - \#(\mathcal{N}(v) \cap A_t) \leq \Delta - d_t(v).$$

The vertices in $\mathcal{N}(v) \cap D_t$ can contribute at most $2 \times \#(\mathcal{N}(v) \cap D_t)$ distinct colors to $K_X(v) \cup K_Y(v)$. The vertices in $\mathcal{N}(v) \cap A_t$ can contribute at most $d_t(v)$ distinct colors. Together they contribute at most $2 \times (\Delta - d_t(v)) + d_t(v) \leq 2\Delta - d_t(v)$ colors.

Many of you argued that the vertices in $\mathcal{N}(v) \cap A_t$ all carry colors that appear in $K_X(v) \cap K_Y(v)$. However, there is no guarantee that each vertex in $\mathcal{N}(v) \cap A_t$ has a distinct color. In the extreme case, both X and Y could use a single color for every vertex in $\mathcal{N}(v) \cap A_t$. The number of colors used for vertices in $\mathcal{N}(v) \cap A_t$ need not equal $d_t(v) = \#(\mathcal{N}(v) \cap A_t)$.

(iii) (5 points) Deduce that

$$\mathbb{P}(f_{t+1} = f_t - 1 \mid \text{info}_t) \geq \alpha f_t + \frac{m_t}{nq}$$

Again condition on V_{t+1}. If $V_{t+1} = v \in D_t$ and $C_{t+1} \notin K_X(v) \cup K_Y(v)$ then both X and Y accept the new color, creating one fewer disagreement.

Thus

$$\begin{aligned}
\mathbb{P}(f_{t+1} &= f_t - 1 \mid \text{info}_t) \\
&\geq \mathbb{P}\{V \in D_t, C \notin K_X(V) \cup K_Y(V) \mid \text{info}_t\} \\
&= \sum\nolimits_{v \in D_t} \mathbb{P}\{V = v \mid \text{info}_t\} \mathbb{P}\{C \notin K_X(v) \cup K_Y(v) \mid \text{info}_t, V = v\} \\
&\geq n^{-1} \sum\nolimits_{v \in D_t} \left(1 - \frac{\#(K_X(v) \cup K_Y(v))}{q}\right) \\
&\geq \frac{1}{n} \sum\nolimits_{v \in D_t} \left(1 - \frac{2\Delta - d_t(v)}{q}\right) \\
&\geq \frac{1}{n}\left(1 - \frac{2\Delta}{q}\right) f_t + \frac{m_t}{nq} \qquad \text{because } f_t = \#D_t \\
&= \alpha f_t + \frac{m_t}{nq}
\end{aligned}$$

Some of you failed to condition properly, treating V (or v) as random in one line and fixed in the next, with summations over $v \in D_t$ appearing out of nowhere. Some of you also confused $\mathbb{I}\{C \notin K_X(v) \cup K_Y(v))\}$ with $\#\left(K_X(v) \cup K_Y(v)\right)$.

[8] Combine the last two results.

(i) (5 points) Show that

$$\begin{aligned}
\mathbb{E}\left(f_{t+1} - f_t \mid \text{info}_t\right) &= \mathbb{P}\{f_{t+1} = 1 + f_t \mid \text{info}_t\} - \mathbb{P}(f_{t+1} = f_t - 1 \mid \text{info}_t) \\
&\leq -\alpha f_t
\end{aligned}$$

From [6](v) and [7](iii),

$$\mathbb{P}\{f_{t+1} = 1 + f_t \mid \text{info}_t\} - \mathbb{P}(f_{t+1} = f_t - 1 \mid \text{info}_t) \leq \frac{m_t}{nq} - \left(\alpha f_t + \frac{m_t}{nq}\right)$$

(ii) (5 points) Deduce that $\mathbb{E}f_{t+1} \leq (1-\alpha)\mathbb{E}f_t$ for all t.

Multiply both sides of $\mathbb{E}\left(f_{t+1} - f_t \mid \text{info}_t\right) \leq -\alpha f_t$ by $\mathbb{P}\{\text{info}_t\}$ then sum over all possible values for info_t to deduce that $\mathbb{E}(f_{t+1} - f_t) \leq -\alpha \mathbb{E} f_t$. Some of you failed to distinguish between $\mathbb{E}\left(f_{t+1} - f_t \mid \text{info}_t\right)$ and $\mathbb{E}\left(f_{t+1} - f_t\right)$.

(iii) (5 points) Deduce that

$$\mathbb{E}f_t \leq (1-\alpha)^t \mathbb{E}f_0 \leq n(1-\alpha)^t.$$

Notice that the upper bound does not depend on μ.

Use $f_0 \leq \#\mathcal{V} = n$ together with recursive appeals to the bound from (ii).

(iv) (5 points) Conclude, via Problems [2] and [3] that

$$\mathbb{TV}_t \leq n(1-\alpha)^t$$

Remark. *You could solve to find how large t must be in order to make $\mathbb{TV}_t \leq \epsilon$, for any given $\epsilon > 0$.*